

STIC-Biotech/ChemLib

88592

From: Schnizer, Richard
Sent: Monday, March 10, 2003 12:11 PM
To: STIC-Biotech/ChemLib
Subject: 09/836,439

Please search the commercial databases for SEQ ID NOS: 1-6 from 09/836,439.

Thank you-

Richard Schnizer, Ph.D.
Patent Examiner
Art Unit 1635
CM1 12E17
703-306-5441
Mail Box CM1 11E12

03/12

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 3/11
Date Completed: 3/19
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on:      March 17, 2003, 09:43:42 ; Search time 625.892 seconds
```

3161.870 Million cell updates/sec

Title: US-09-836-439-1

Sequence: 1 ccttccaacctagtagcag.....ggaagggcgcgttttcgcgc 68

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processed for: Minimum Match On

Maximum Match 100%
Listing first 45 summaries

Database :

1:	gb.ba:*
2:	gb.htg.*
3:	gb.un.*
4:	gb.om.*
5:	gb.ov.*
6:	gb.pat.*
7:	gb.ph.*
8:	gb.pl.*
9:	gb.pr.*
10:	gb.ro.*
11:	gb.sts.*
12:	gb.sy.*
13:	gb.un.*
14:	gb.v1.*
15:	gb.ba.*
16:	em.fun.*
17:	em.hum.*
18:	em.ln.*
19:	em.mu.*
20:	em.om.*
21:	em.ov.*
22:	em.ov.*
23:	em.pat.*
24:	em.ph.*
25:	em.pl.*
26:	em.ro.*
27:	em.sts.*
28:	em.un.*
29:	em.v1.*
30:	em.hum.*
31:	em.hlg.lnv.*
32:	em.hlg.other.*
33:	em.hlg.mus.*
34:	em.hlg.pln.*
35:	em.hlg.tod.*
36:	em.hlg.mem.*
37:	em.hlg.vrt.*
38:	em.sv.*
39:	em.hgo.hum.*
40:	em.hgo.mus.*
41:	em.hgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	28	41.2	68	6	AR024440	AR024440 Sequence
2	28	41.2	68	6	193657	193657 Sequence 23
3	27.2	40.0	10591	2	AC020132	AC020132 Drosophila
4	27.2	40.0	73076	3	AC003055	AC003055 Drosophila
5	27.2	40.0	168540	3	AC092189	AC092189 Drosophila
6	27.2	40.0	303356	3	AE003583	AE003583 Drosophila
7	27	39.7	3554	3	DME721	X00854 Drosophila
8	27	39.7	56043	2	AC012761	AC012761 Drosophila
9	27	39.7	66991	3	AC001653	AC001653 Drosophila
10	27	39.7	74026	2	AC101196	AC101196 Mus muscu
11	27	39.7	80866	2	AC012649	AC012649 Drosophi
12	27	39.7	110000	3	AE001572_1	Contigene (2 of
13	27	39.7	190642	3	AC095015	Drosophila
14	27	39.7	208375	2	AC107703	AC107703 Mus muscu
15	27	39.7	309357	3	AE003673	Drosophila
16	26.6	39.1	91104	2	AC112654	AC113664 Rattus no
17	26.6	39.1	160589	10	AL663066	Rattus no
18	26.4	38.8	68	6	AR024439	Sequence
19	26.4	38.8	68	6	193656	Sequence 22
20	26.4	38.8	20281	1	AE008861	AE008861 Salmonell
21	26.4	38.8	192675	10	AL589767	Mouse DNA
22	26.4	38.8	199842	2	AC094068	Rattus no
23	26.4	38.8	207922	10	AL450399	Mouse DNA
24	26.2	38.5	139877	9	AC004066	Mouse DNA
25	26.2	38.5	194521	2	AC093696	Homo sapl
26	26	38.2	161179	2	AC109157	Homo sapl
27	26	38.2	183311	2	AC109788	Mus muscu
28	26	38.2	187352	10	AL607109	Mouse DNA
29	26	38.2	222941	2	AC102236	Mus muscu
30	25.8	37.9	2693	10	MMPB	X60133 Murine MPB
31	25.8	37.9	2703	3	MMPBLT	X87952 Murine MPB
32	25.8	37.9	2743	10	MMPDE	X55668 Mouse mRNa
33	25.8	37.9	129790	3	AG441131	AJ441131 Anopheles
34	25.8	37.9	157418	2	AC006912	AC006912 Caenorhab
35	25.8	37.9	163475	9	AC092619	AC092619 Homo sapl
36	25.6	37.6	154552	9	AL390964	AL390964 Human DNA
37	25.6	37.6	190721	8	AP003263	AP003263 Oryza sat
38	25.4	37.4	435	6	AX341469	AX341469 Sequence
39	25.4	37.4	1224	9	BC007520	BC007520 Homo sapl
40	25.4	37.4	1324	9	AF026222	AF026222 Homo sapl
41	25.4	37.4	2727	6	AX282833	AX282833 Sequence
42	25.4	37.4	2730	6	AX282845	AX282845 Sequence
43	25.4	37.4	2874	6	AX282837	AX282837 Sequence
44	25.4	37.4	2877	6	AX282849	AX282849 Sequence
45	25.4	37.4	2879	6	AX282841	AX282841 Sequence

ALIGNMENTS

RESULT	1			
AR024440				
LOCUS	AR024440	68 bp	DNA	linear
DEFINITION	Sequence 23 from patent US 5795972.			PAT 05-DEC-1998
ACCESSION	AR024440			
VERSION	AR024440.1	GI:3977734		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
	Unclassified.			
REFERENCE	1 (bases 1 to 68)			
AUTHORS	Kmiec,E.B.			
TITLE	Chimeric mutational vectors having non-natural nucleotides			
JOURNAL	Patent: US 5795972-A 23-18-ANG-1998.			
FEATURES	Location/Qualifiers			

```

source      1. .68
            /organism="unknown"
BASE COUNT  11 a 19 c 19 g 19 t
ORIGIN
Query Match      41.2%; Score 28; DB 6; Length 68;
Best Local Similarity 61.7%; Pred. No. 4.3;
Matches 37; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

OY 9 CCTACGTAGCAGAAAGTTTACUUCUCUACGTAGCGUUGAAGGCGCGTTTCCGCG 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    CCTGAGGAGAGAGACTGCTTTGCGAGTCTCTCTCAGAGAGTCAAGTCCGCGTTTCCGCG 68

RESULT 2
193657 LOCUS      193657 68 bp DNA linear PAT 01-DEC-1998
DEFINITION      Sequence 23 from patent US 5731181.
ACCESSION      193657
VERSION      193657.1 GI:3938127
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 68)
AUTHORS      Kmiec, E.B.
TITLE      Chimeric mutational vectors having non-natural nucleotides
JOURNAL      Patent: US 5731181-A 23-24-MAR-1998;
FEATURES
SOURCE      1. .68
            /organism="unknown"
BASE COUNT  11 a 19 c 19 g 19 t
ORIGIN
Query Match      41.2%; Score 28; DB 6; Length 68;
Best Local Similarity 61.7%; Pred. No. 4.3;
Matches 37; Conservative 3; Mismatches 20; Indels 0; Gaps 0;

OY 9 CCTACGTAGCAGAAAGTTTACUUCUCUACGTAGCGUUGAAGGCGCGTTTCCGCG 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    CCTGAGGAGAGAGACTGCTTTGCGAGTCTCTCTCAGAGAGTCAAGTCCGCGTTTCCGCG 68

Db 9 CCTGAGGAGAGAGACTGCTTTGCGAGTCTCTCTCAGAGAGTCAAGTCCGCGTTTCCGCG 68

RESULT 3
AC020132/c AC020132 10591 bp DNA linear HTG 03-JAN-2000
LOCUS      Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION      pieces.
ACCESSION      AC020132
VERSION      AC020132.1 GI:6664765
KEYWORDS      HTG; HTGS; PHASE2.
SOURCE      Drosophila melanogaster.
ORGANISM      Drosophila melanogaster.
REFERENCE      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
TITLE      Ephydriidae; Drosophilidae; Drosophila.
JOURNAL      1 (bases 1 to 10591)
COMMENT      Adams, M. and Venter, J.C.
            Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
            This sequence was identified as CDN:10212153 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            Location/Qualifiers
            1. 10591
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
BASE COUNT  2920 a 2216 c 2177 g 3278 t
ORIGIN

```

```

Query Match      40.0%; Score 27.2; DB 2; Length 10591;
Best Local Similarity 60.7%; Pred. No. 12;
Matches 34; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

OY 11 TACGTAGCAGAAAGTTTACUUCUCUACGTAGCGUUGAAGGCGCGTTTCCGCG 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    TACGTAGCAGAAAGTTTACUUCUCUACGTAGCGUUGAAGGCGCGTTTCCGCG 66

Db 3352 TACGTAGCAGAAAGTTTACUUCUCUACGTAGCGUUGAAGGCGCGTTTCCGCG 66

RESULT 4
AC003055 AC003055 73076 bp DNA linear INV 04-NOV-1997
LOCUS      Drosophila melanogaster (PI DS06332 (D91)) DNA sequence, complete
DEFINITION      sequence.
ACCESSION      AC003055
VERSION      AC002945 AC001994 AC002947 AC002944 AC002952 AC001992
KEYWORDS      AC002946 AC001988 AC001995 AC002951 AC001991 AC002842
            AC001993 AC002948 AC002953 AC001990 AC001989 AC002949 AC002950
            AC003055.1 GI:2584828
SOURCE      HTG.
ORGANISM      Drosophila melanogaster (Subclones in pot2 from PI clone DS06332
            (D91)) DNA.
REFERENCE      Drosophila melanogaster
AUTHORS      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydriidae; Drosophilidae; Drosophila.
            1 (bases 1 to 73076)
            Ceiliker, S.E., Aghavan, A., Arcaina, T.T., Baxter, E., Doyle, C.M.,
            Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K.,
            Kearney, L., Kim, S.H., Ko, C.L., Li, M., Lomocan, M.A., Mazda, P.,
            Mok, M.S., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, D.,
            Santos, R.F., Snit, E., Stevko, V., Subramanian, S., Towne, B.,
            Wan, K.H., White, K.R., Yee, A., Zhang, R., Zieran, L.L. and
            Kimmel, B.
            Sequencing of Drosophila chromosome 2L, region 22F
            Unpublished (1997)
            2 (bases 1 to 73076)
            Ceiliker, S.E., Aghavan, A., Arcaina, T.T., Baxter, E., Doyle, C.M.,
            Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K.,
            Kearney, L., Kim, S.H., Ko, C.L., Li, M., Lomocan, M.A., Mazda, P.,
            Mok, M.S., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, D.,
            Santos, R.F., Snit, E., Stevko, V., Subramanian, S., Towne, B.,
            Wan, K.H., White, K.R., Yee, A., Zhang, R., Zieran, L.L. and
            Kimmel, B.
            Direct Submission
            Submitted (04-NOV-1997) Berkeley Drosophila Genome Project, MS
            74-157, Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, US
            Sequence submitted by:
            Lawrence Berkeley National Laboratory, MS 74-157
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.hgc.lbl.gov/sequence-archive.html) or
            send email to drosophila@genome.lbl.gov.
            Library location: 66-92.
            This PI was assembled from the following subclones: 2_c1
            (AC001994), 2_b12 (AC002947), 1_g6 (AC002944), 1_d4 (AC002943),
            2_b5, 2_e4 (AC002952), 1_b4, 2_a1, 2_a2, 1_b6 (AC001992), 1_h8
            (AC002945), 2_a8 (AC002946), 1_b2 (AC001988), 2_c5 (AC001995), 2_d5
            (AC002951), 1_h3 (AC001991), 1_d3 (AC002942), 2_a9 (AC001993),
            2_c10 (AC002948), 2_g9 (AC002953), 1_g7 (AC001990), 1_b5
            (AC001989), 2_c3 (AC002949), 2_h2, 2_c4 (AC002950), 2_b6.
            Location/Qualifiers
            1. 73076
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /map="22F2-22F4"
            /clone="PI DS06332 (D91)"
BASE COUNT  21203 a 14890 c 15769 g 21214 t
ORIGIN

```


misc_feature	/note="poc. alternate translation start site for ftz protein".1927	
intron	1778..1927	
exon	/number=1 1928..2867	
misc_feature	/number=2 1933..2121	
misc_feature	/note="region with homology to Antp and Ubx genes" 2023..2121	
variation	/note="poc. DNA-binding domain of the ftz protein" 2041..2043	
variation	/note="GCC (Ala) is GTC (Val) in temperature sensitive allele ftz (f47ts)" 2112..2142	
variation	/note="mutation ftz (Rpl) causing 10 novel amino acids and a premature stop codon (2140-2142)" 2844..2849	
polyA_signal	/note="poc. polyadenylation signal" 2848..2854	
polyA_signal	/note="poc. polyadenylation signal" 2857..2867	
polyA_signal	/note="poc. polyadenylation signal" 979 a 952 c 831 g 792 t	
BASE COUNT	979 a 952 c 831 g 792 t	
ORIGIN		
Query Match	39.7%; Score 27; DB 3; Length 354;	
Best Local Similarity	58.8%; Pred. No. 14;	
Matches	30; Conservative 6; Mismatches 15; Indels 0; Gaps 0;	
QY	15 TAGCAGAAAGTTTACUUCUGCAGCTAGCGUUGAAGCGCGTTTCG 65	
Db	171 TTGCATTAAGTTTACTGTTACTAGTACATTGGGAAGTCGTTGTGG 221	
RESULT 8		
AC012761		
LOCUS	AC012761	
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.	
ACCESSION	AC012761	
VERSION	AC012761.1 GI:6223082	
KEYWORDS	HTG: HTGS_PHASE2.	
SOURCE	Drosophila melanogaster.	
ORGANISM	Drosophila melanogaster.	
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
AUTHORS	1 (bases 1 to 56043)	
TITLE	Adams,M. and Venter,J.C.	
JOURNAL	Direct Submission	
COMMENT	Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA	
	This sequence was identified as CDM:10210082 by the submitter. For further information on this sequence you may e-mail to fly@celera.com.	
	* NOTE: This is a 'working draft' sequence.	
	* This sequence will be replaced	
	* by the finished sequence as soon as it is available and	
	* the accession number will be preserved.	
	Location/Qualifiers	
FEATURES	1..56043	
source	/organism="Drosophila melanogaster"	
BASE COUNT	16447 a 11385 c 11452 g 16759 t	
ORIGIN		
Query Match	39.7%; Score 27; DB 2; Length 56043;	
Best Local Similarity	58.8%; Pred. No. 17;	

Matches 30: Conservative 6: Mismatches 15: Indels 0: Gaps 0:

QY 15 TAGCAGAAAGTTTACUUCGUCAGTAGUGAAGGCGGCTTTTCG 65

Db 5971 TTGCATTAAGTTTACTGTTACTACTGATCTTTGGAAGTGCCTTTG 6021

RESULT 9
AC001653 66991 bp DNA linear INV 16-APR-1999

LOCUS Drosophila melanogaster, chromosome 3R, region 84B1-84B2, P1 clone

DEFINITION D507876, complete sequence.

ACCESSION AC001653 L49396 L39779 L32657 L32652 L32645 L32656 L32649 L32648

VERSION L32637 L39749 L32654 L32639 L32647 L32642 L32635 L32655 L32651

KEYWORDS L32638 L32655 L32640 L32641 L32646 L32650 L32636 L32643 L32644

SOURCE AC001653.1 GI:2342708

ORGANISM HTG.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

REFERENCE Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

AUTHORS 1 (bases 1 to 66991)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazer, R.G.,

Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,

Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,

Syrskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

TITLE Sequencing of antennapedia complex, homeotic genes

JOURNAL Unpublished (1998)

REFERENCE 2 (bases 1 to 66991)

AUTHORS Martin, C.H., Arcaina, T.T., Bondoc, M.M., Chiang, A., Critz, P.A.,
Davis, C.A., Doyle, C.M., Ericsson, C.L., Farfan, D.E., Gunning, K.M.,

Houston, K.A., Jaklevic, M.A., Kadner, K.E., Kim, K., Kim, S.F.,
Ko, C.L., Lewis, K.D., Li, M., Lindquist, K.J., Lomotan, M.A.,

Lustre, V.M., Machrus, M.U., Mayeda, C.A., Miguel, T.M., Miller, C.A.,
Mok, M.S., Pacled, J.M., Patel, S.G., Santos, R.F., Sudamanian, S.,

Wan, K.H., Whitelaw, K.R., Yee, A., Yeh, R.T., Yu, C. and Palazzolo, M.J.

TITLE Direct Submission

JOURNAL Submitted (22-APR-1997) Berkeley Drosophila Genome Project, MS

COMMENT 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US

On or before Apr 16, 1999 this sequence version replaced gi:483988,
gi:483986, gi:1945589, gi:1103946.

Sequence submitted by:
Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.

P1 library location: 83-4.

FEATURES Location/Qualifiers

1..66991

/organism="Drosophila melanogaster"

/strain="y2; cn bw sp"

/db_xref="taxon:7227"

/chromosome="3R"

/map="84B1-84B2"

/clone="P1 D507876 (D14)"

/note="This sequence has not changed since its original
submission on 08/25/1997. It was resubmitted in order to
include all secondary accession numbers for the subclones
belonging to this clone."

BASE COUNT 19565 a 13581 c 13571 g 20274 t

Query Match 39.7% Score 27; DB 3; Length 66991;

Best Local Similarity 58.8%; Pred. No. 17;

Matches 30: Conservative 6: Mismatches 15: Indels 0: Gaps 0:

QY 15 TAGCAGAAAGTTTACUUCGUCAGTAGUGAAGGCGGCTTTTCG 65

Db 25751 TTGCATTAAGTTTACTGTTACTACTGATCTTTGGAAGTGCCTTTG 25801

RESULT 10
AC101196 74026 bp DNA linear HTG 23-NOV-2001

LOCUS Mus musculus clone RP23-177G5, LOW-PASS SEQUENCE SAMPLING.

DEFINITION AC101196

ACCESSION AC101196.1 GI:17059970

VERSION HTG; HTGS_PHASE0.

KEYWORDS Mus musculus.

SOURCE Mus musculus.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 74026)

Birren, B., Linton, L., Nussbaum, C. and Lander, E.

Unpublished

2 (bases 1 to 74026)

AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,

Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Chopel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
Lamazzares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,

Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Melidim, J.,

Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Center project name: 116139

Center clone name: 177_G5

NOTE: This record contains 93 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

RESULT 11	AC012649	80866 bp	DNA	linear	HTG 04-NOV-1999
LOCUS	AC012649				
DEFINITION	Drosophila melanogaster chromosome X clone BACR0708 (D1121) RPI-98 07.N.8 map 17D-17E strain Y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 101 unordered pieces.				
ACCESSION	AC012649				
VERSION	AC012649.2				
KEYWORDS	HTG; HTG-PHASE1.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 80866) Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Keatney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Richards, S., Sethi, H., Svirkas, R.R., Wan, K.H., Webster, D., Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.				
TITLE	Sequencing of Drosophila melanogaster				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 80866) Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Keatney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomolan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.				
COMMENT	Direct Submission Submitted (02-NOV-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA On Nov 4, 1999 this sequence version replaced g1:6175114. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to dbgap@fruitfly.berkeley.edu . All contigs in this submission meet the following cutoffs: length >= 200 bases. * NOTE: this is a 'working draft' sequence. It currently * consists of 101 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.				
	1	577: contig of 577 bp in length			
	578	657: gap of unknown length			
	658	1404: contig of 747 bp in length			
	1405	1484: gap of unknown length			
	1485	2313: contig of 829 bp in length			
	2314	2393: gap of unknown length			
	2394	3131: contig of 738 bp in length			
	3132	3211: gap of unknown length			
	3212	3719: contig of 508 bp in length			
	3720	3799: gap of unknown length			
	3800	4897: contig of 1098 bp in length			
	4898	4977: gap of unknown length			
	4978	5799: contig of 322 bp in length			
	5800	5879: gap of unknown length			
	5880	6469: contig of 590 bp in length			
	6470	6549: gap of unknown length			
	6550	7442: contig of 893 bp in length			
	7443	7532: gap of unknown length			
	7523	8486: contig of 964 bp in length			
			8487	8566: gap of unknown length	
			8567	9246: contig of 680 bp in length	
			9247	9326: gap of unknown length	
			9327	9371: contig of 645 bp in length	
			9372	9971: gap of unknown length	
			9972	10051: gap of 581 bp in length	
			10052	10632: contig of 581 bp in length	
			10633	10712: gap of unknown length	
			10713	11328: contig of 616 bp in length	
			11329	11408: gap of unknown length	
			11409	12070: contig of 662 bp in length	
			12071	12150: gap of unknown length	
			12151	12954: contig of 804 bp in length	
			12955	13034: gap of unknown length	
			13035	13887: contig of 853 bp in length	
			13888	13967: gap of unknown length	
			13968	14675: contig of 708 bp in length	
			14676	14755: gap of unknown length	
			14756	15777: contig of 1022 bp in length	
			15778	15857: gap of unknown length	
			15858	16561: contig of 704 bp in length	
			16562	16641: gap of unknown length	
			16642	17656: contig of 1015 bp in length	
			17657	17736: gap of unknown length	
			17737	18347: contig of 611 bp in length	
			18348	18427: gap of unknown length	
			18428	19269: contig of 842 bp in length	
			19270	19349: gap of unknown length	
			19350	20123: contig of 774 bp in length	
			20124	20203: gap of unknown length	
			20204	21037: contig of 834 bp in length	
			21038	21117: gap of unknown length	
			21118	21860: contig of 763 bp in length	
			21861	21960: gap of unknown length	
			21961	23063: contig of 1103 bp in length	
			23064	23143: gap of unknown length	
			23144	23742: contig of 599 bp in length	
			23743	23822: gap of unknown length	
			23823	24697: contig of 875 bp in length	
			24698	24777: gap of unknown length	
			24778	25699: contig of 922 bp in length	
			25700	25779: gap of unknown length	
			25780	26547: contig of 768 bp in length	
			26548	26627: gap of unknown length	
			26628	27682: contig of 1055 bp in length	
			27683	27762: gap of unknown length	
			27763	28171: contig of 409 bp in length	
			28172	28251: gap of unknown length	
			28252	28623: contig of 572 bp in length	
			28624	28903: gap of unknown length	
			28904	29965: contig of 1062 bp in length	
			29966	30045: gap of unknown length	
			30046	31173: contig of 1128 bp in length	
			31174	31253: gap of unknown length	
			31254	32156: contig of 903 bp in length	
			32157	32236: gap of unknown length	
			32237	32872: contig of 636 bp in length	
			32873	32952: gap of unknown length	
			32953	33817: contig of 865 bp in length	
			33818	33897: gap of unknown length	
			33898	35066: contig of 1169 bp in length	
			35067	35146: gap of unknown length	
			35147	37168: contig of 2042 bp in length	
			37169	37258: gap of unknown length	
			37259	38406: contig of 1138 bp in length	
			38407	38487: gap of unknown length	
			38487	40075: gap of 1589 bp in length	
			40076	40155: gap of unknown length	
			40156	41204: contig of 1049 bp in length	
			41205	41284: gap of unknown length	
			41285	42613: contig of 1329 bp in length	
			42614	42693: gap of unknown length	
			42694	45350: contig of 2657 bp in length	
			45351	45430: gap of unknown length	

```

* 45431 45914: contig of 484 bp in length
* 45915 45994: gap of unknown length
* 45995 45989: contig of 595 bp in length
* 46590 46669: gap of unknown length
* 46670 47222: contig of 553 bp in length
* 47223 47302: gap of unknown length
* 47303 47887: contig of 585 bp in length
* 47888 47968: gap of unknown length
* 47968 48554: contig of 586 bp in length
* 48554 48634: gap of unknown length
* 48634 49267: contig of 633 bp in length
* 49267 49347: gap of unknown length
* 49347 49944: contig of 598 bp in length
* 49945 50024: gap of unknown length
* 50025 50642: contig of 618 bp in length
* 50643 50722: gap of unknown length
* 50723 51227: contig of 505 bp in length
* 51228 51307: gap of unknown length
* 51308 52034: contig of 727 bp in length
* 52035 52114: gap of unknown length
* 52115 52686: contig of 572 bp in length
* 52687 53363: gap of unknown length
* 53363 53443: contig of 597 bp in length
* 53444 53993: gap of unknown length
* 53994 54073: contig of 350 bp in length
* 54074 54434: gap of unknown length
* 54435 55129: contig of 361 bp in length
* 55129 55209: gap of unknown length
* 55210 55887: contig of 615 bp in length
* 55888 55967: gap of unknown length
* 55968 56510: contig of 678 bp in length
* 56511 56590: contig of 543 bp in length
* 56591 57022: gap of unknown length
* 57023 57102: contig of 432 bp in length
* 57103 57394: gap of unknown length
* 57395 57474: contig of 292 bp in length
* 57475 58106: gap of unknown length
* 58107 58186: contig of 632 bp in length
* 58187 58920: gap of unknown length
* 58921 59000: contig of 734 bp in length
* 59001 59609: gap of unknown length
* 59610 59689: contig of 609 bp in length
* 59690 60371: gap of unknown length
* 60372 60451: contig of 682 bp in length
* 60452 61001: gap of unknown length
* 61002 61081: contig of 550 bp in length
* 61082 61656: gap of unknown length
* 61657 61736: contig of 575 bp in length
* 61737 62261: gap of unknown length
* 62262 62341: contig of 525 bp in length
* 62342 62865: gap of unknown length
* 62866 62945: contig of 524 bp in length
* 62946 63538: gap of unknown length
* 63539 63618: contig of 593 bp in length
* 63619 64136: gap of unknown length
* 64137 64216: contig of 518 bp in length
* 64216: gap of unknown length

```

```

Query Match
Best Local Similarity 39.7%; Score 27; DB 2; Length 80866;
Matches 30; Conservative 6; Mismatches 15; Indels 0; Gaps 0;
Oy 15 TAGCAGAAAGTTTACUUCUGACAGTACGAGGCGCGTTTTCG 65
Db 25511 TTGCATAAAGTTTACTGTTTACTAGTCAATTTGGAAGTGCCTTGTGG 25561

```

```

RESULT 12
AE001572-1/c
WPCOMMENT
Sequence split into 5 fragments
Fragment Name Begin End

```

```

AE001572-0 1 110000
AE001572-1 100001 210000
AE001572-2 200001 310000
AE001572-3 300001 410000
AE001572-4 400001 429825
Continuation (2 of 5) of AE001572 from base 100001 (AE001572 Drosophila melanogaster)

Query Match
Best Local Similarity 39.7%; Score 27; DB 3; Length 110000;
Matches 30; Conservative 6; Mismatches 15; Indels 0; Gaps 0;
Oy 15 TAGCAGAAAGTTTACUUCUGACAGTACGAGGCGCGTTTTCG 65
Db 93289 TTGCATAAAGTTTACTGTTTACTAGTCAATTTGGAAGTGCCTTGTGG 93239

```

```

RESULT 13
AC095015 190642 bp DNA linear INV 15-SEP-2001
DEFINITION BACR32J03, complete sequence.
LOCUS AC095015
VERSION AC095015
KEYWORDS AC095015.1 GI:15624857
SOURCE HTG.
ORGANISM Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 190642)
Celinker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,J.F., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresner,D., Farfan,D.,
Fierlater,S., Fris,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
Pacheb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 84A-84B
Unpublished
2 (bases 1 to 190642)

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Celinker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,J.F., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresner,D., Farfan,D.,
Fierlater,S., Fris,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
Pacheb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

```

```

TITLE
JOURNAL
COMMENT
Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

```

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES
Source Location/Qualifiers
1.190642
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="84A-84B"
/clone="BACR32J03 (D1350)"
/clone_id="RPCT-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcotI in
pBACE3.6)"

BASE COUNT 55773 a 39290 c 38978 g 56601 t

ORIGIN
Query Match 39.7%; Score 27; DB 3; Length 190642;
Best Local Similarity 58.8%; Pred. No. 18;
Matches 30; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

OY 15 TAGCAGAAAGTTTACUUCGACGACGAGUUGAGAGCGCGTTTCG 65
Db 30503 TTGCATTAAGTTTACTGTTACTAGTCATTTTGGAAGTGCCTTTGTTGG 30553

RESULT 14
AC107703 208375 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone RP23-472H1, WORKING DRAFT SEQUENCE, 8 ordered
DEFINITION pieces.
AC107703
AC107703.5 GI:22381033
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FUZZTOP.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 208375)
Birren, B., Nusbaum, C., and Lander, E.
Mus musculus, clone RP23-472H1
Unpublished
2 (bases 1 to 208375)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
Landers, T., Lehoczek, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McGowan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Rett, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stefayes, S., Theodore, J.,
Strauss, N., Sudramanlian, A., Talamas, J., Testafaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Tsigillo, O., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

TITLE
JOURNAL Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 208375)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepe, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,

TITLE
JOURNAL Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:21326325.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L18630
Center clone name: 472.H-1
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 206755 bases at least Q40
Consensus quality: 207261 bases at least Q30
Consensus quality: 207461 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 207675; sum-of-contigs
Quality coverage: 11.1 in Q20 bases; sum-of-contigs
Quality coverage: 10.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 53805: contig of 53805 bp in length
53806 53905: gap of 100 bp
53906 56169: contig of 2264 bp in length
56170 56269: gap of 100 bp
56270 62030: contig of 5761 bp in length
62031 62130: gap of 100 bp
62131 71647: contig of 9517 bp in length
71648 71747: gap of 100 bp
71748 84166: contig of 12419 bp in length
84167 84266: gap of 100 bp
84267 138240: contig of 53974 bp in length
138241 138340: gap of 100 bp
138341 186299: contig of 47959 bp in length
186300 186399: gap of 100 bp
186400 208375: contig of 21976 bp in length.

FEATURES
Source Location/Qualifiers
1.208375
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-472H1"
/clone_lib="RPCT-23 female Mouse BAC"
1.53805
/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature
misc_feature
misc_feature
53906.56169

TITLE
JOURNAL
MEDLINE
20196006

PUBMED
10731132

REFERENCE
2 (bases 1 to 309357)

AUTHORS
Adams,M.D., Celisnker,S.E., Glibbs,R.A., Rubin,G.M. and Venter,C.J.

JOURNAL
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

COMMENT
On Oct 9, 2000 this sequence version replaced gi:7298860.
location/Qualifiers

FEATURES
source

gene

1. 309357
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3R"
13397..>24361
/gene="Dfd4"
/note="CG2189"
/map="84A5-84A5"
/db_xref="FLYBASE:FBan0002189"
/db_xref="FLYBASE:FBgn0000439"
join(13397..14933,15319..15419,15485..15662,22801..23187,
23445..24361)
/gene="Dfd"
/product="C7y132"

CDS

/note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."
/db_xref="FLYBASE:FBan0002189"
/db_xref="FLYBASE:FBgn0000439"
join(14147..14933,15319..15419,15485..15662,22801..23187,
23445..23761)
/gene="Dfd"

/note="Dfd gene product; Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."

/codon_start=1
/protein_id="AA554083.1"
/db_xref="GI:7298876"
/db_xref="FLYBASE:FBan0002189"
/db_xref="FLYBASE:FBgn0000439"
/translation="MSSPLMGYPAPHANPHVQSMSCNGCLDPKFPFLADDDYHNHYGNHS
MTATSGMSGAGVGACGVSVCGGCGAGCTGPRHMPRADMSDTMAHHNPNSHS
HTHSIPRHNSNAISCHHOASAGTSSNTANKTPSPRHSHPAHPQSLSLYVNHP
EFTSAGVSDPDPTNGTGPAPANVPNTSSNGGGSGLAVLGAGVGSANGYGCGGY
GOOLGVSGTHSQSHSPSQMDLPLOCSTPSPNYALGLELGLKERLEAEVA
ETTDGEITILYPMWKIHIVAGVANGSYOGEMERKQRRTAYTRHOTILEKEPNRYRLT
RRRIETLAHTLVLESEROIKITFONRMKKWCKDKLRPKTKRKRTVDANGKPVAK
PKATRASSKSQQOAOOOQSOQOQTOTPTMYEVSCITSDSLSGDVSSSLGPPYI
PAAPTTSPIPSQOHLNNNNNGSNNNNNNNNNNNMLNNNNNNOMGHNLHGHILO
QOOSDLTNLDLHKIDDYDTAL"

gene

complement(<46894..>70354)
/gene="Scr"
/note="CG1030"
/map="84A5-84B1"
/db_xref="FLYBASE:FBan0001030"
/db_xref="FLYBASE:FBgn0003339"
complement(join(46894..47918,63464..64443,70145..70354))
/gene="Scr"
/product="Cr1096"

/note="Nucleotide sequence of the Celera sequence differs from the published sequence for this transcript."

/db_xref="FLYBASE:FBan0001030"
/db_xref="FLYBASE:FBgn0003339"
complement(join(47611..47918,63464..64409))
/gene="Scr"

/note="Scr gene product; Nucleotide sequence of the Celera

XX	FT		/note="DNA nucleotides"
PN	PB		
XX	PD	M09741141-AI.	
XX	PE	06-NOV-1997.	
XX	PF	01-MAY-1997;	9/WO-US07362.
PR	PG	01-MAY-1996;	96US-0640517.
XX	PH	(UYJE-) UNIV JEFFERSON THOMAS.	
PI	PI	Cole-strauss A, Kmiec EB, Yoon K;	
XX	PP	WPI; 1997-549675/50.	
PT	PT	Chimeric nucleic acid repair vectors - used for treating diseases such as sickle cell disease, beta-thalassemia, Gaucher disease, hypercholesterolaemia, emphysema or haemophilia	
PS	PS	Disclosure; Fig 3; 79pp: English.	
XX	XX	This chimeric repair vector (CRV) SC4 is used in comparative studies on the experimental use of a CRV SCI designed to repair the mutation found in sickle cell disease beta-globin and the beta-globin of a HSC. The CRV designed to repair the mutation contains a nucleic acid having at most one 3' end and one 5' end comprising a segment of unpaired bases disposed. The unpaired bases separate the nucleic acid into a first strand and a second strand, comprising a first region and a second region respectively, each region having at least 15 nucleotides. Each nucleotide of the first region is Watson-Crick paired to a nucleotide of the second region and the first region comprises at least 8 ribonucleotides, which are Watson-Crick paired to 2'-deoxynucleotides, which ribonucleotides form at least one ribonucleotide segment of at least 3 ribonucleotides and the sequence of the first or the second region is the sequence of a fragment of a wild-type allele of a human gene. The CRVs can be used for repairing genetic mutations in cells for re-introducing into a patient for treating diseases. They can be used for treating, sickle cell disease, beta-thalassemia, familial hypercholesterolaemia, Gaucher disease, emphysema or haemophilia.	
CC	CC	Sequence 68 BP; 11 A; 19 C; 19 G; 13 T; 6 U; 0 other;	
SQ	SQ		
Query Match		41.2%; Score 28; DB 18; Length 68;	
Best Local Similarity		66.7%; Pred. No. 0.57;	
Matches 40; Conservative		0; Mismatches 20; Indels 0; Gaps 0	
Oy		9 CCTACGAGCAGCAATTTCACUUCUGCUCGACGAGGUGGAGGGCGTTTGCAGC 68 Db 9 CTCGAGGAGAAGACTCTTTTGACAGUCUUCUCCCTGAGGAGCAGGUGCGGCTTTTCCGCG 68	
RESULT 2			
AAVI2904			
DD	AAVI2904 standard; DNA; 68 BP.		
AC	AAVI2904;		
XX	XX		
DT	17-JUN-1998 (first entry)		
XX	XX		
DE	Chimeric mutational vector SC4.		
KM	Chimeric mutational vector; alkaline phosphatase gene; gene repair;		
KW	disease-related mutation; human; Gaucher's disease; sickle-cell anaemia;		
KW	thalassaemia; familial hypercholesterolaemia; emphysema; therapy;		
KW	circular; ss.		
OS	Synthetic.		
XX	XX		
Key	Location/Qualifiers		
FT	misc_feature	1..25	
FT	/*tag=	a	
FT	/note="binds to nucleotides 54 to 30"		

FT		misc_feature	30..54	/tag= b
PT			/note= "binds to nucleotides 25 to 1"	
FT		misc_RNA	30..39	/tag= c
FT		misc_RNA	45..54	/tag= d
FT		misc_feature	55..59	/tag= e
FT			/note= "binds to nucleotides 68 to 64"	
FT		misc_feature	64..68	/tag= f
FT			/note= "binds to nucleotides 59 to 55"	
XX				
PN		M09748714-A1.		
XX				
PD		24-DEC-1997.		
XX				
PF		16-JUN-1997;	97MO-US10538.	
XX				
PR		17-JUN-1996;	96US-0664487.	
XX				
PA		(UYJE-) UNIV JEFFERSON THOMAS.		
XX				
Kmlec EB;				
XX				
DR		WPI; 1998-063068/06.		
XX				
XX		Oligonucleotide for altering a genomic sequence in eukaryotes - particularly for correcting disease-related mutation(s) and for production of transgenic animals and plants		
XX				
PS		Example 7.2; Fig 3; 68bp; English.		
XX				
CC		This sequence represents a nucleotide analogue of the invention, termed a chimeric mutational vector (CMV). This sequence is directed against the beta-globin gene. The CMVs (I) are for altering a gene in a eukaryotic cell, and comprise: (a) a first strand (S1) having at least 15 nucleotides (nt); at least 3 nuclease-resistant ribo-type nt (nr') and at least 3 contiguous ribo-type nt, the same as, or additional to, the nr'; and (b) a second strand (S2) in which the nt are Watson-Crick (WC) paired to the nt in S1. The contiguous ribo-type nt in S1 are WC-paired to 2'-deoxyribo-type nt, and at least one ribo-type nt is other than a 2'-O-methyl substituted nt. nr' are 2'-AX-nucleosides; 2'-AX-nucleosides or 2'-AR-nucleotides; A = oxygen, fluoro, chloro or bromo; when A = O, X = hydrogen or 1-6C alkyl and R = 1-6C alkyl; when A is halo then R and X are absent. (I) are used to repair a disease-related mutation in human cells (e.g. those associated with Gaucher's disease, sickle-cell anaemia, thalassemia, familial hypercholesterolaemia, emphysema etc.). The chimeric mutation vectors are also used to inactivate specific genes, i.e. to generate transgenic ('knockout') animals or plants. They are also CC used for biomedical research and for pharmaceutical production. Any CC eukaryotic gene of known sequence can be altered, by replacement, deletion or addition.		
CC				
XX		Sequence 68 BP; 11 A; 19 C; 19 G; 13 T; 6 U; 0 other:		
SQ				
		Query Match	41.2%; Score 28; DB 19; Length 68;	
		Best Local Similarity	66.7%; Pred. No. 0.57;	
		Matches	40; Conservative	0; Mismatches
			20; Indels	0; Gaps
YY		9 CCTACGTAGCAGAAGATTTTTACUUDUCGCACTAGCGTUGGAAGGGCGGCTTTGCGCC 68 Dbg 9 C CTGAGAGAGAAGACTCTTTTCAGUCUUCUCCCTAGAGACGACGAGUCGCGCTTTGCGCC 68		
RESULT 3				
AAAX19657				
ID		AAAX19657 standard; DNA; 68 BP.		
XX				
AC		AAAX19657;		
XX				
YT		02-JUN-1999 (first entry)		

[illegible]

XX	FH	Key	Location/Qualifiers
XX	FT	misc-feature	26..29
XX	FT	/tag=	a
XX	FT	/note=	"nucleotides replacing pentaethylene oxide (PEO)
XX	FT		loop portion"
XX	PN		
XX	PD	MO9839353-AI.	
XX	PD	11-SEP-1998.	
XX	PF	19-FEB-1998;	98WO-US03224A.
XX	PR	03-MAR-1997;	97US-0039244.
XX	PA	(PEKE) PERKIN-ELMER CORP.	
XX	PI	Andrus A. Kulmellis RG;	
XX	DR	WPI; 1998-495787/42.	
XX	PT	New chimeric oligo:nucleotide(s) - which are useful for causing	
XX	PT	specific alterations in target nucleic acids	
XX	PS	Example 3; Page 33; 56pp; English.	
CC	CC	This oligonucleotide is used in the course of the invention. The	
CC	CC	invention provides chimeric oligonucleotides comprising a stem portion	
CC	CC	having one or more nucleotides selected from 2'-O-alkyl-ribonucleotides,	
CC	CC	2'-O-allyl-ribonucleotides, 2'-allyl-ribonucleotides, 2'-halo-	
CC	CC	ribonucleotides, 2'-O-methoxyethyl-ribonucleotides, 2'-branching group-	
CC	CC	ribonucleotides or 2'-O-branching group-ribonucleotides, and one or more	
CC	CC	loop portions. The chimeric oligonucleotides may be used for introduction	
CC	CC	of specific alterations into target nucleic acid sequences located in	
CC	CC	living organisms. They have increased intramolecular and intermolecular	
CC	CC	duplex stability, increased resistance to nuclease degradation and	
CC	CC	increased chemical stability, resistance to hydrolysis and degradation,	
CC	CC	as compared to known chimeric oligonucleotides. They have structures	
CC	CC	which induce intramolecular and intermolecular A-type helix formation.	
CC	CC	They have highly stable conformations, as measured by thermal melting	
CC	CC	measurements.	
SQ	SQ	Sequence 68 BP; 13 A; 17 C; 17 G; 21 T; 0 other;	
XY	XY	Query Match	39.7%; Score 27; DB 19; Length 68;
XY	XY	Best Local Similarity	58.2%; Pred. No. 1.3;
XY	XY	Matches 39; Conservative	3; Mismatches 25; Indels 0; Gaps 0
OY	OY	2 CTTCACACCCTAGCTACGAGAAGTTTTCACUUCUGCAGCTAGGUGGAAGGCCGCTT	61
OY	OY		
OY	OY		
OY	OY	2 CTTCACAAGGATGCTACTATTATTTTAAGTAGAGCATCCCTGTGGAAAGTGCGGCTT	61
OY	OY		
OY	OY		
OY	OY	62 TTCGCGC	68
OY	OY		
OY	OY	62 TTGC GCGC	68
ABLO3968/c	ABLO3968/c	standard; CDNA: 3785 BP.	
ABL03968;	ABL03968;		
DT	DT	26-MAR-2002 (first entry)	
DE	DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 6386.	
KW	KW	Drosophila: developmental biology; cell signalling; insecticide;	
KW	KW	pharmaceutical; gene; ss.	
OS	OS	Drosophila melanogaster.	
NN	NN	WO200171042-A2.	

XX	27-SEP-2001.
PD	
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI: 2001-656860/75.
DR	P-PADB; ABB59865.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT	
PS	Claim 1; SEQ ID NO 6386; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins (ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	Sequence 3785 BP; 876 A; 851 C; 982 G; 1076 T; 0 other;
QY	
Dn	Query Match 39.7%; Score 27; DB 23; Length 3785; Best Local Similarity 58.8%; Pred. No. 3.3; Matches 30; Conservative 6; Mismatches 15; Indels 0; Gaps 0. 15 TAGCAGAAAGTTTTCACUUCUGCUACGATGAGUGGAGCGCGCTTTCG 65 : : : : : 3541 TTGCATTAAAGTTTTCGTCTTACTGTTACTAGCATTTTGGAAGTCGTTGGTGG 3491
RESULT 6	
AAV09395	AAV09395 standard; DNA; 68 BP.
AC	AAV09395;
XX	
DT	14-MAY-1998 (first entry)
XX	
DE	Chimeric repair vector (CRV) SC3.
OS	Synthetic.
OS	Homo sapiens.
XX	
FH	Key
FT	misc_feature Location/Qualifiers
FT	1..29
FT	/tag= a
FT	/note= "DNA nucleotides"
FT	30..39
FT	/tag= b
FT	/note= "RNA nucleotides"
FT	40..44
FT	/tag= c
FT	/note= "DNA nucleotides"
FT	45..54
FT	/tag= d
FT	misc_RNA

FT		/note= "RNA nucleotides"
FT	misc_feature	55..68
FT		/*tag= e
XX		/note= "DNA nucleotides"
FN	MO974114I-AI.	
PD	06-NOV-1997.	
XX		
XX	01-MAY-1997;	97WO-USO7362.
PF		
PR	01-MAY-1996;	96US-0640517.
XX		
PA	(UYJE-) UNIV JEFFERSON THOMAS.	
XX		
PI	Cole-strauss A, Kmiec EB, Yoon K;	
DR	WPI; 1997-549675/50.	
XX		
PT	Chimeric nucleic acid repair vectors - used for treating diseases such as sickle cell disease, beta-thalassemia, Gaucher disease, hypercholesterolaemia, emphysema or haemophilia	
PT		
PS	Disclosure; Fig 3; 79pp; English.	
XX		
CC	This chimeric repair vector (CRV) SC3 is used in comparative studies on the experimental use of a CRV SC1 designed to repair the mutation found in sickle cell disease beta-globin and the beta-globin of a HSC. The CRV designed to repair the mutation contains a nucleic acid having at most one 3' end and one 5' end comprising a segment of unpaired bases disposed. The unpaired bases separate the nucleic acid into a first strand and a second strand, comprising a first region and a second region respectively, each region having at least 15 nucleotides. Each nucleotide of the first region is Watson-Crick paired to a nucleotide of the second region and the first region comprises at least 8 ribonucleotides, which are Watson-Crick paired to 2'-deoxynucleotides, which ribonucleotides form at least one ribonucleotide segment of at least 3 ribonucleotides and the sequence of the first or the second region is the sequence of a fragment of a wild-type allele of a human gene. The CRVs can be used for repairing genetic mutations in cells for re-introducing into a patient for treating diseases. They can be used for treating, sickle cell disease, beta-thalassemia, familial hypercholesterolaemia, Gaucher disease, emphysema or haemophilia.	
CC		
XX		
SQ	Sequence 68 BP; 12 A; 18 C; 18 G; 14 T; 6 U; 0 other;	
	Query Match	38.8%; Score 26.4; DB 18; Length 68;
	Best Local Similarity	65.0%; Pred. No. 2.2;
	Matches 39; Conservative	0; Mismatches 21; Indels 0; Gaps 0;
OY	9 CCTRACGTAGCAGAAGATTTTTACUUCUUGCUAGCTAGGUGGAAGCGCGTTTGCGGC 68 Db 9 CCTGAGAGAGAAAGACTCTTTTCACUUCUUCUCCCTAGAGAGUACAAGCGCGTTTGCGGC 68	
RESULT 7		
AAV12903		
ID	AAV12903 standard; DNA; 68 BP.	
XX		
AC	AAV12903;	
XX		
DT	17-JUN-1998 (first entry)	
XX		
DE	Chimeric mutational vector SC3.	
XX		
KW	Chimeric mutational vector; alkaline phosphatase gene; gene repair; disease-related mutation; human; Gaucher's disease; sickle-cell anaemia; thalassemia; familial hypercholesterolaemia; emphysema; therapy; circular; ss.	
KW		
OS	Synthetic.	
XX		
PH	Key Location/Qualifiers	

FT	misc_feature	1..25
FT	/tag=	a
PT	/note=	"binds to nucleotides 54 to 30"
FT	misc_feature	30..54
FT	/tag=	b
FT	/note=	"binds to nucleotides 25 to 1"
FT	misc_RNA	30..39
FT	/tag=	c
FT	misc_RNA	45..54
FT	/tag=	d
FT	misc_feature	55..59
FT	/tag=	e
FT	/note=	"binds to nucleotides 68 to 64"
FT	misc_feature	64..68
FT	/tag=	f
FT	/note=	"binds to nucleotides 59 to 55"
XX		
PN	WO97A8714-A1.	
PD	24-DEC-1997.	
PF	16-JUN-1997;	97WQ-USI0538.
PR	17-JUN-1996;	96US-0664487.
PA	(UYJE-) UNIV JEFFERSON THOMAS.	
PI	kmlcc EB;	
DR	WPI; 1998-063068/06.	
PT	Oligonucleotide for altering a genomic sequence in eukaryotes - particularly for correcting disease-related mutation(s) and for production of transgenic animals and plants	
PS	Example 7.2; Fig 3; 68pp; English.	
XX	This sequence represents a nucleotide analogue of the invention, termed a chimeric mutational vector (CMV). This sequence is directed against the beta-globin gene. The CMVs (I) are for altering a gene in a eukaryotic cell, and comprise: (a) a first strand (S1) having at least 15 contiguous nt; at least 3 nuclease-resistant ribo-type nt ('nt') and at least 3 contiguous ribo-type nt, the same as, or additional to, the nt'; and (b) a second strand (S2) in which the nt are Watson-Crick (WC) paired to the nt in S1. The contiguous ribo-type nt in S1 are WC-paired to 2'-deoxyribo-type nt, and at least one ribo-type nt is other than a 2'-O-methyl substituted nt. nt' are 2'-AX-nucleosides; 2'-AX-nucleosides or 2'-AR-nucleotides: A = oxygen, fluoro, chloro or bromo; when A = O, X = hydrogen or 1-6c alkyl and R = 1-6c alkyl; when A is halo then R and X are absent. (I) are used to repair a disease-related mutation in human cells (e.g. those associated with Gaucher's disease, sickle-cell anaemia, thalassemia, familial hypercholesterolaemia, emphysema etc.). The chimeric mutation vectors are also used to inactivate specific genes, i.e. to generate transgenic ('knockout') animals or plants. They are also used for biomedical research and for pharmaceutical production. Any eukaryotic gene of known sequence can be altered, by replacement, deletion or addition.	
SQ	Sequence 68 BP; 12 A; 18 C; 18 G; 14 T; 6 U; 0 other:	
	Query Match	38.8%; Score 26.4; DB 19; Length 68;
	Best Local Similarity	65.0%; Pred. No. 2.2;
	Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;	
Gy	9 CCTAGTAGCAAAATTTCACUUUUCGCUACGTAGUGGAGAAGCGCTTTTGGCCG 68 Db 9 CCTGAGGAGAACATCGCTTTTGCAGUUCUCCTCAGGAGUACAUGCCTTTTGCACC 68	
RESULT 8		
ID	AAX19656	
XX	AAX19656 standard; DNA; 68 BP.	

AC AAXI9656;
XX
DT 02-JUN-1999 (first entry)
XX Oligonucleotide SEQ ID NO:95.
DE
KW Genome; genetic lesion; haematopoietic stem cell; hepatocyte; RNase;
KM human wild-type allele; mutation; sickle cell anaemia; thalassemia;
KV Gaucher's disease; glucocerebrosidase gene; hypercholesterolaemia;
WM emphysema; haemophilia; Christmas disease; ss.
XX
OS Synthetic.
PN US588983-A.
PD 30-MAR-1999.
XX
PF 05-AUG-1997; 97US-0906265.
XX
PR 05-AUG-1997; 97US-0906265.
PR 01-MAY-1996; 96US-0640517.
PR 17-JUN-1996; 96US-0664487.
XX
PA (UYJE-) UNIV JEFFERSON THOMAS.
PI Cole-Strauss AD, Kmiec EB:
DR WPI: 1999-243264/20.
XX
PT Double-stranded oligonucleotides with containing a human wild-type
PT allele - useful for repairing mutations in human cells,
PT particularly those causing sickle cell anaemia or thalassemia
XX
PS Disclosure: Column 59-60; 40pp: English.

The present invention describes double-stranded oligonucleotides (I) containing fragments of wild-type human alleles. (II) are used to repair disease associated mutations in human cells. (I) are preferably used to treat sickle cell anaemia or thalassemia (mutations in the beta-globin gene, including the promoter region), or Gaucher's disease (mutations in the glucocerebrosidase gene), In haematopoietic cells. (I) may also be used to treat familial hypercholesterolemia (mutations in the low-density lipoprotein receptor gene), emphysema (the alpha 1-anti-trypsin gene), haemophilia (the factor VIII gene) or Christmas disease (the factor IX gene), in hepatoerythrocytic cells. (I) provides repair of small genetic mutations. The present sequence represents an oligonucleotide from the present invention.

Sequence 68 BP; 12 A; 18 C; 18 G; 14 T; 6 U; 0 other:

Query Match 38.8%; Score 26.4; DB 20; Length 68;
Best Local Similarity 65.0%; Pred.No.2.2;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

DY 9 CCTAGTAGCAAAAGTTTTCACUUCGCGCAGGAGGGCGCTTGCGGC 68
||| | |||| | |||| | |||| | ||||
ID 9 CCTGAGAGAACAATGTTCAGACUCCUCCCTCAGGAGUCAGAUGGCCGTTTTCGCGC 68

RESULT 9
ABQ51034
ID ABQ51034 standard; DNA: 533 BP.
XX
AC ABQ51034;
XX
DT 12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 37625.

Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KM gastrointestinal; respiratory system; single nucleotide polymorphism;
SNP; cell differentiation; ds.

XX Homo sapiens.
 OS WO200218632-A2.
 PN 07-MAR-2002.
 PD 01-SEP-2001; 2001WO-EP10074.
 PF 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX (EPiG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K, Guetig D;
 PI WPI; 2002-371829/40.
 DR
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 PS Claim 12; 56pp; German.
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 CC
 XX
 SQ Sequence 533 BP; 89 A; 62 C; 208 G; 174 T; 0 other;
 Query Match 37.6%; Score 25.6; DB 24; Length 533;
 Best Local Similarity 55.4%; Pred. No. 7.1;
 Matches 31; Conservative 6; Mismatches 19; Indels 0; Gaps 0;
 QY 12 ACGTAGCAGAAAGTTTACUUCUGACGTAGGUGGAGGCGGCTTTTCGCG 67
 Db 1 ATGTCCGGCGGATTTTATTATTATTTAGTAGAAGGTGGAAGCGGAGATTAGGG 56
 RESULT 10
 AB051035/C
 ID AB051035 standard; DNA; 533 BP.
 XX
 AC AB051035;
 XX
 DT 12-JUL-2002 (first entry)
 DE
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 37626.
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX

PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP10074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX 05-SEP-2000; 2000DE-1044543.
 XX (EPiG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K, Guetig D;
 PI WPI; 2002-371829/40.
 DR
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 PS Claim 12; 56pp; German.
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 CC
 XX
 SQ Sequence 533 BP; 174 A; 208 C; 62 G; 89 T; 0 other;
 Query Match 37.6%; Score 25.6; DB 24; Length 533;
 Best Local Similarity 55.4%; Pred. No. 7.1;
 Matches 31; Conservative 6; Mismatches 19; Indels 0; Gaps 0;
 QY 12 ACGTAGCAGAAAGTTTACUUCUGACGTAGGUGGAGGCGGCTTTTCGCG 67
 Db 533 ATGTCCGGCGGATTTTATTATTATTTAGTAGAAGGTGGAAGCGGAGATTAGGG 478
 RESULT 11
 ABL38127
 ID ABL38127 standard; CDNA; 435 BP.
 XX
 AC ABL38127;
 XX
 DT 08-APR-2002 (first entry)
 DE
 XX Human colon tumour antigen polynucleotide SEQ ID NO:1716.
 KW Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
 KW colon tumour metastatic antigen; diagnosis; gene; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO200196388-A2.
 XX
 PD 20-DEC-2001.
 XX
 PF 08-JUN-2001; 2001WO-US18557.

PR	07-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217487
PR	11-JUL-2000	2000US-0217496
PR	11-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225113
PR	14-AUG-2000	2000US-0225114
PR	14-AUG-2000	2000US-0225667
PR	14-AUG-2000	2000US-0225676
PR	14-AUG-2000	2000US-0225678
PR	14-AUG-2000	2000US-0225750
PR	14-AUG-2000	2000US-0225847
PR	14-AUG-2000	2000US-0225870
PR	14-AUG-2000	2000US-0225877
PR	14-AUG-2000	2000US-0225944
PR	14-AUG-2000	2000US-0225958
PR	14-AUG-2000	2000US-0225958
PR	14-AUG-2000	2000US-0225979
PR	18-AUG-2000	2000US-0226579
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	05-SEP-2000	2000US-0229305
PR	05-SEP-2000	2000US-0229309
PR	05-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0230437
PR	08-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232387
PR	14-SEP-2000	2000US-0232387
PR	14-SEP-2000	2000US-0232388
PR	14-SEP-2000	2000US-0232389
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234273
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235854
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236127
PR	29-SEP-2000	2000US-0236167
PR	29-SEP-2000	2000US-0236168
PR	29-SEP-2000	2000US-0236169
PR	29-SEP-2000	2000US-0236170
PR	02-OCT-2000	2000US-0236602
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	13-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241807
PR	20-OCT-2000	2000US-0241808

CC diseases (e.g. cardiac insufficiency, coronary insufficiency or high
CC blood pressure). The GNG DNA and protein sequences of the invention may
CC also be used as insulin sensitizers - for improving insulin sensitivity
CC in persons with non-insulin dependent diabetes mellitus. The present cDNA
CC sequence encodes the human GNG-7A protein.
XX

SO Sequence 2257 BP; 728 A; 522 C; 482 G; 525 T; 0 other;

Query Match 37.4%; Score 25.4; DB 24; Length 2257;
Best Local Similarity 55.9%; Pred. No. 12;
Matches 33; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

QY 2 CTTCGACCTACGAGAGAACTTTTACUUCGUCGAGGUGGAGGCGCGT 60
DB 1849 CTTCGACCACTTGAAGAACTTTGATCTTGCAGCTTGATTTAATGGCCAGT 1907

RESULT 14
AAK52451

ID AAK52451 standard; cDNA; 2454 BP.

AC AAK52451;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 1980.

KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

PE 05-FEB-2001; 2001WO-US04098.

PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AJ, Yang Y, Wejhrman T, Goodrich R;

WPI: 2001-476283/51.
P-PDB: AAM79318.

Nucleic acids encoding polypeptides with cytokine-like activities,
useful in diagnosis and gene therapy -

Claim 1: Page 4386-4387; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC Inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX

SO Sequence 2454 BP; 782 A; 574 C; 532 G; 563 T; 3 other;

Query Match 37.4%; Score 25.4; DB 22; Length 2454;
Best Local Similarity 55.9%; Pred. No. 12;
Matches 33; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

QY 2 CTTCGACCTACGAGAGAACTTTTACUUCGUCGAGGUGGAGGCGCGT 60
DB 2175 CTTCGACCACTTGAAGAACTTTGATCTTGCAGCTTGATTTAATGGCCAGT 2233

RESULT 15
AAK52452

ID AAK52452 standard; cDNA; 2454 BP.

AC AAK52452;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 1981.

KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

PE 05-FEB-2001; 2001WO-US04098.

PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AJ, Yang Y, Wejhrman T, Goodrich R;

WPI: 2001-476283/51.
P-PDB: AAM79319.

Nucleic acids encoding polypeptides with cytokine-like activities,
useful in diagnosis and gene therapy -

Claim 1: Page 4387-4388; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

Tue Mar 18 16:16:11 2003

us-09-836-439-1.rng

Page 10

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX Sequence 2454 BP; 782 A; 574 C; 532 G; 563 T; 3 other;
SQ

Query Match	37.4%;	Score 25.4;	DB 22;	Length 2454;
Best Local Similarity	55.9%;	Pred. No. 15;		
Matches 33;	Conservative	5;	Mismatches 21;	Indels 0;
				Gaps 0;

Oy 2 CTTCAACCTAGCTAGCAGAAAATTGTTTTACUUCUCUACGTGATGSGUGGAAGGGGCGGT 60
 |||| | | | | | | | | | | : | | : | | : | |
Db 2175 CTTCGAAACAATTAGAGAATCTTTCATCTTCAGCTGTGATTAATTATGCCCACT 2233

Search completed: March 17, 2003, 10:50:28
Job time : 393.887 secs


```

source
1. .818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT      218 a      197 c      181 g      222 t
ORIGIN

Query Match      39.7% Score 27; DB 12; Length 818;
Best Local Similarity 57.6% Pred. No. 17;
Matches 34; Conservative 5; Mismatches 20; Indels 0; Gaps 0;

2 CTTCCAACTTACGTGACGAAGTTTTCUUCUUCGACUACGTAGGUGGAGGCGCG 60
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
466 CTTTCGACCACTTAGGAGAACCTTTGATCTTCACGCTTGATGATTTAATGGCAGT 522

RESULT 2
LOCUS CNS00341
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR07N08 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL063844
AL063844.1 GI:4941600
GSS.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org/TheBDGP/Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2: cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR07N08"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT      349 a      179 c      216 g      325 t      32 others
ORIGIN

Query Match      39.7% Score 27; DB 17; Length 1101;
Best Local Similarity 58.8% Pred. No. 19;
Matches 30; Conservative 6; Mismatches 15; Indels 0; Gaps 0;

15 TAGCAGAAAGTTTTCUUCUUCGACUACGTAGGUGGAGGCGCGTTTCG 65

```

DB	693	TTGCATTAAGTTTCTTACTGTTACTGATGATTTTGGAACTGGCGTTTGTGG	743
RESULT 3			
LOCUS	B1739136/c		
DEFINITION	B1739136	741 bp	mRNA
ACCESSION	603361404F1 NIH_MGC_94	Mus musculus	cdna clone IMAGE:5368368 5',
VERSION	RNA sequence.		
KEYWORDS	B1739136		
SOURCE	B1739136.1	GI:15716149	
ORGANISM	EST.		
	house mouse.		
	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Euteleia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 741)		
JOURNAL	NIH-MGC http://mgc.ncl.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-remail.nih.gov		
	Tissue Procurement: The Cepko Laboratory		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: L14M11937 row: h column: 01		
	High quality sequence stop: 729.		
FEATURES	Location/Qualifiers		
Source	1..741		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:5368368"		
	/clone_lib="NIH_MGC_94"		
	/tissue_type="retina"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: eye; Vector: pCMW-SPOrt6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally; oligo-dT primed.		
	Average insert size 3.3 kb. Library enriched for		
	full-length clones and constructed by Life Technologies.		
	Note: this is a NIH-MGC Library."		
BASE COUNT	210 a	177 c	194 g 160 t
ORIGIN			
	Query Match	37.9%	Score 25.8; DB 13; Length 741;
	Best Local Similarity	69.0%;	Pred. No. 47;
	Matches 20; Conservative	7; Mismatches	2; Indels 0; Gaps 0;
QY	30	ACUUUCUGCUACGAGUGUGAGAGGCCG	58
	: : : :		
Db	368	ACTTCTGCTACGATGAGTGGAGGCCAC	340
RESULT 4			
LOCUS	B1733738/c	747 bp	mRNA
DEFINITION	60335355F1 NIH_MGC_94	Mus musculus	cdna clone IMAGE:5359908 5',
ACCESSION	RNA sequence.		
VERSION	B1733738		
KEYWORDS	B1733738.1	GI:15710751	
SOURCE	EST.		
ORGANISM	house mouse.		
	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Euteleia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 747)		
TITLE	NIH-MGC http://mgc.ncl.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		

Email: c9apbs-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: LLM13915 row: g column: 13
 High quality sequence stop: 734.

FEATURES

Location/Qualifiers
 1. 747

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:5359908"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT

218 a 169 c 198 g 162 t

ORIGIN

Query Match

Best Local Similarity 37.9%; Score 25.8; DB 13; Length 747;
 69.0%; Pred. No. 47;
 Matches 20; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Qy

30 ACUUCUGCAGTAGGUGGAGGCGC 58

Db

183 ACTTCTGCTACGTAGTGAAGCCAC 155

RESULT 5

BQ946406 899 bp mRNA linear EST 21-AUG-2002

LOCUS

BQ946406/8924053 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6467512

DEFINITION

5', mRNA sequence.

ACCESSION

BQ946406

VERSION

BQ946406.1 GI:22361884

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE

NIH-MGC http://mgc.nci.nih.gov/.

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

FEATURES

Tissue Procurement: The Cepko Laboratory

source

cDNA Library Preparation: Life Technologies, Inc.

FEATURES

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

AUTHORS

DNA Sequencing by: Agencourt Bioscience Corporation

TITLE

Clone distribution: MGC clone distribution information can be

JOURNAL

found through the I.M.A.G.E. Consortium/LNLN at:

COMMENT

http://image.llnl.gov
 Plate: LLM13993 row: 1 column: 17
 High quality sequence stop: 614.

Full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT

248 a 208 c 231 g 207 t 5 others

ORIGIN

Query Match

Best Local Similarity 37.9%; Score 25.8; DB 14; Length 899;
 69.0%; Pred. No. 50;
 Matches 20; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Qy

30 ACUUCUGCAGTAGGUGGAGGCGC 58

Db

532 ACTTCTGCTACGTAGTGAAGCCAC 504

RESULT 6

BQ937919 923 bp mRNA linear EST 21-AUG-2002

LOCUS

BQ937919/8932027 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6465857

DEFINITION

5', mRNA sequence.

ACCESSION

BQ937919

VERSION

BQ937919.1 GI:22353397

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE

NIH-MGC http://mgc.nci.nih.gov/.

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

FEATURES

Tissue Procurement: The Cepko Laboratory

source

cDNA Library Preparation: Life Technologies, Inc.

FEATURES

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

AUTHORS

DNA Sequencing by: Agencourt Bioscience Corporation

TITLE

Clone distribution: MGC clone distribution information can be

JOURNAL

found through the I.M.A.G.E. Consortium/LNLN at:

COMMENT

http://image.llnl.gov
 Plate: LLM13989 row: d column: 18
 High quality sequence stop: 666.

FEATURES

Location/Qualifiers
 1. 923

source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:6465857"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT

251 a 224 c 233 g 214 t 1 others

ORIGIN

Query Match

Query Match

Best Local Similarity 37.9%; Score 25.8; DB 14; Length 923;
 69.0%; Pred. No. 50;
 Matches 20; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Qy

30 ACUUCUGCAGTAGGUGGAGGCGC 58

Db

667 ACTTCTGCTACGTAGTGAAGCCAC 639

RESULT 7

BG262012 943 bp mRNA linear EST 13-FEB-2001

LOCUS

BG262012/60233784F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4481291 5'

DEFINITION

mRNA sequence.

ACCESSION

BG262012

VERSION

BG262012.1 GI:12771828

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 943)
NIH-MGC <http://mgc.ncl.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10316 row: f column: 12
High quality sequence stop: 694.
Location/Qualifiers
1. 943
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4481943"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORE6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 269 a 230 c 256 g 188 t
ORIGIN

Query Match 37.9%; Score 25.6; DB 12; Length 943;
Best Local Similarity 69.0%; Pred. No. 51;
Matches 20; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 30 ACUUCUGCUCAGTAGGUGGAGGCGC 58
||||:||||:||||:||||:||||:|
Db 471 ACTTCTGCTAGCTAGGTGGAGGCCAC 443

RESULT 8
BG342737/c 961 bp mRNA linear EST 27-FEB-2001
LOCUS 602374557F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:4481943 5',
DEFINITION mRNA sequence.
ACCESSION BG342737
VERSION BG342737.1 GI:13149175
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 961)
NIH-MGC <http://mgc.ncl.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10318 row: a column: 16
High quality sequence stop: 653.
Location/Qualifiers

source 1. 961
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4481943"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORE6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 288 a 218 c 254 g 201 t
ORIGIN

Query Match 37.9%; Score 25.8; DB 12; Length 961;
Best Local Similarity 69.0%; Pred. No. 51;
Matches 20; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 30 ACUUCUGCUCAGTAGGUGGAGGCGC 58
||||:||||:||||:||||:||||:|
Db 31 ACTTCTGCTAGCTAGGTGGAGGCCAC 3

RESULT 9
BG295440/c 1101 bp mRNA linear EST 21-FEB-2001
LOCUS 602392754F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:4504476 5',
DEFINITION mRNA sequence.
ACCESSION BG295440
VERSION BG295440.1 GI:13057077
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1101)
NIH-MGC <http://mgc.ncl.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10376 row: 1 column: 13
High quality sequence stop: 735.
Location/Qualifiers
1. 1101
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:4504476"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORE6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 329 a 279 c 289 g 204 t
ORIGIN

Query Match 37.9%; Score 25.8; DB 12; Length 1101;
Best Local Similarity 69.0%; Pred. No. 54;
Matches 20; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 30 ACUUCUGCUCAGTAGGUGGAGGCGC 58
||||:||||:||||:||||:||||:|

Db 260 ACGTACGTAGCAGCATATTATTAGTGTAGTACGTACGTACGATGG 213

RESULT 13
A0070730 349 bp mRNA linear EST 02-APR-2002

LOCUS A0070730 Rice cDNA from young root Oryza sativa (japonica cultivar-group) cDNA clone R10161_2A, mRNA sequence.

ACCESSION A0070730.1 GI:5038620

VERSION EST.
KEYWORDS Oryza sativa (japonica cultivar-group).
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 349)
AUTHORS Yamamoto, K. and Sasaki, T.
TITLE Rice cDNA from young root
JOURNAL Unpublished (1999)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/PROJECT="RGP"

FEATURES
source Location/Qualifiers

1..349
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="R10161_2A"
/clone_lib="Rice cDNA from young root"
/issue_type="young root"
BASE COUNT 58 a 98 c 96 g 97 t
ORIGIN

Query Match 37.6%; Score 25.6; DB 9; Length 349;
Best Local Similarity 60.4%; Pred. No. 43;
Matches 29; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 8 ACGTACGTAGCAGCAAGTTTACUUCUGCAGTACGTACGAGGAGG 55
Db 85 ACGTACGTAGCAGCATATTATTAGTGTAGTACGTACGATGG 132

RESULT 14
A0057142/c 352 bp mRNA linear EST 01-APR-2002
LOCUS A0057142 Oryza sativa mature leaf Nipponbare Oryza sativa (japonica cultivar-group) cDNA clone S21188_1A, mRNA sequence.
ACCESSION A0057142
VERSION A0057142.1 GI:4716026
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Oryza sativa (japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 352)
AUTHORS Yamamoto, K. and Sasaki, T.
TITLE Rice cDNA from mature leaf
JOURNAL Unpublished (1999)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/

FEATURES
source Location/Qualifiers

1..352
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="S21188_1A"
/clone_lib="Oryza sativa mature leaf Nipponbare"
/issue_type="mature leaf"
BASE COUNT 97 a 97 c 100 g 58 t
ORIGIN

Query Match 37.6%; Score 25.6; DB 9; Length 352;
Best Local Similarity 60.4%; Pred. No. 44;
Matches 29; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 8 ACGTACGTAGCAGCAAGTTTACUUCUGCAGTACGAGGAGG 55
Db 268 ACGTACGTAGCAGCATATTATTAGTGTAGTACGTACGATGG 221

RESULT 15
A0162789/c 356 bp mRNA linear EST 03-APR-2002
LOCUS A0162789 Rice mature leaf Oryza sativa (japonica cultivar-group) cDNA clone S21904, mRNA sequence.
ACCESSION A0162789
VERSION A0162789.1 GI:11026188

KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Oryza sativa (japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 356)
AUTHORS Sasaki, T. and Yamamoto, K.
TITLE Rice cDNA from mature leaf (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/PROJECT="RGP"

FEATURES
source Location/Qualifiers

1..356
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="S21904"
/clone_lib="Rice mature leaf"
/issue_type="mature leaf"
BASE COUNT 98 a 99 c 99 g 58 t 2 others
ORIGIN

Query Match 37.6%; Score 25.6; DB 9; Length 356;
Best Local Similarity 60.4%; Pred. No. 44;
Matches 29; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 8 ACGTACGTAGCAGCAAGTTTACUUCUGCAGTACGAGGAGG 55
Db 272 ACGTACGTAGCAGCATATTATTAGTGTAGTACGTACGATGG 225

Search completed: March 17, 2003, 13:09:07
Job time : 2409.76 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 230.108 Seconds
(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-2

Perfect score: 25
Sequence: 1 ccatgacacattggaatgagag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pal:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_man:*
37: em_htg_vtc:*
38: em_sy:*
39: em_htg_hum:*
40: em_htg_mus:*
41: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.4	93.6	85	10	MMHIFLAS04
2	23.4	93.6	538	6	AF004144 Mus muscu
3	23.4	93.6	2481	6	AF004144 Mus muscu
4	23.4	93.6	2481	6	AF004144 Mus muscu
5	23.4	93.6	2509	6	AF004144 Mus muscu
6	23.4	93.6	2522	6	AF004144 Mus muscu
7	23.4	93.6	2528	6	AF004144 Mus muscu
8	23.4	93.6	2530	6	AF004144 Mus muscu
9	23.4	93.6	2537	6	AF004144 Mus muscu
10	23.4	93.6	2711	10	MMHIFLAS04
11	23.4	93.6	3551	9	AF004144 Mus muscu
12	23.4	93.6	3678	9	AF004144 Mus muscu
13	23.4	93.6	3718	10	AF004144 Mus muscu
14	23.4	93.6	3746	6	AF004144 Mus muscu
15	23.4	93.6	3746	10	MMHIFLAS04
16	23.4	93.6	3867	10	MMHIFLAS04
17	23.4	93.6	3927	6	AF004144 Mus muscu
18	23.4	93.6	3933	9	AF004144 Mus muscu
19	23.4	93.6	3942	9	AF004144 Mus muscu
20	23.4	93.6	3973	10	AF004144 Mus muscu
21	23.4	93.6	4183	10	AF004144 Mus muscu
22	23.4	93.6	10355	6	AF004144 Mus muscu
23	23.4	93.6	20463	10	MMHIFLAS04
24	23.4	93.6	169792	2	AC104313
25	23.4	93.6	188107	9	CNS01DME
26	23.4	93.6	228786	2	AC124712
27	21.8	87.2	2551	4	AB018398
28	20.4	81.6	171838	2	AC105321
29	20.2	80.8	3006	5	AF02782
30	20.2	80.8	3076	5	AF212989
31	20.2	80.8	117771	9	AC112906
32	20.2	80.8	181825	2	AC068191
33	19.8	79.2	141762	2	HS198C21
34	19.8	79.2	14837	2	AC021509
35	19.8	79.2	165840	9	AC068112
36	19.8	79.2	187277	9	AC022968
37	19.8	79.2	200000	2	AP000494
38	19.8	79.2	200567	2	AP001587
39	19.8	79.2	203540	8	AP002396
40	19.4	77.6	1596	8	VSCL30RP
41	19.4	77.6	2110	8	SCYGR14BC
42	19.4	77.6	2679	8	SCYGR149W
43	19.4	77.6	2677	8	SCYGR149W
44	19.4	77.6	201802	2	AC124473
45	19.4	77.6	264009	2	AC115294

ALIGNMENTS

RESULT 1
LOCUS MMHIFLAS04
DEFINITION Mus musculus hypoxia-inducible factor 1 alpha (Hif1a) gene, exon 4.
ACCESSION AF004144
VERSION AF004144.1 GI:2197137
KEYWORDS
SEGMENT
SOURCE
ORGANISM
Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Luo, G., Gu, Y. Z., Jain, S., Chan, W. K., Carr, K. M., Hogenesch, J. B. and
Bradfield, C. A.

TITLE Molecular characterization of the murine Hif-1 alpha locus
JOURNAL Gene Expr. 6 (5), 287-299 (1997)
MEDLINE 98034461
PUBMED 9368100
REFERENCE 2 (bases 1 to 85)
AUTHORS Luo, G., Gu, X.-Z., Jain, S., Chan, W.K., Carr, K.M., Hogenesch, J.B. and Bradfield, C.A.
JOURNAL Submitted (14-MAY-1997) Oncology, University of Wisconsin-Madison, 1400 University Ave., Madison, WI 53706, USA
FEATURES
source
1.85
/organism="Mus musculus"
/db_xref="taxon:10090"
1.85
/gene="Hif1a"
/number=4
BASE COUNT 29 a 14 c 18 g 24 t
ORIGIN
Query Match 93.6%; Score 23.4; DB 10; Length 85;
Best Local Similarity 96.0%; Pred. No. 0.43;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CCATGTGACCATGAGGAATGAGAG 25
Db 40 CCATGTGACCATGAGGAATGAGAG 64
RESULT 2
LOCUS HSHF1A03 538 bp DNA linear PRI 26-OCT-1998
DEFINITION Homo sapiens hypoxia-inducible factor 1 alpha subunit (HIF1A) gene,
exons 3 and 4.
ACCESSION AF050117
VERSION AF050117.1 GI:3790523
KEYWORDS 3 of 13
SEGMENT
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 538)
AUTHORS Iyer, N.V., Leung, S.W. and Semenza, G.L.
TITLE The human hypoxia-inducible factor 1alpha gene: HIF1A structure and evolutionary conservation
JOURNAL Genomics 52 (2), 159-165 (1998)
MEDLINE 99000835
PUBMED 9782081
REFERENCE 2 (bases 1 to 538)
AUTHORS Iyer, N.V., Leung, S.W. and Semenza, G.L.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-1998) Departments of Pediatrics and Medicine, Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 600 N. Wolfe St, Baltimore, MD 21287-3914, USA
FEATURES
source
1.538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q21-q24"
1.67
/gene="HIF1A"
/number=2
68.213
/gene="HIF1A"
/number=3
214.297
/gene="HIF1A"
/number=3
298.382
/gene="HIF1A"
/number=4
exon
intron
exon

intron 383.2538
/gene="HIF1A"
/number=4
BASE COUNT 170 a 69 c 89 g 210 t
ORIGIN
Query Match 93.6%; Score 23.4; DB 9; Length 538;
Best Local Similarity 96.0%; Pred. No. 0.41;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CCATGTGACCATGAGGAATGAGAG 25
Db 337 CCATGTGACCATGAGGAATGAGAG 361
RESULT 3
LOCUS AX451938 2481 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0212326.
ACCESSION AX451938
VERSION AX451938.1 GI:21698761
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Poellinger, L., Pereira, T. and Ruas, J.
TITLE Mechanism of conditional regulation of the hypoxia-inducible factor-1 by the von hippel-Lindau tumor suppressor protein
JOURNAL Patent: WO 0212326-A 3 14-FEB-2002;
Angiogenetics Sweden AB (SE)
FEATURES
source
1.2481
/organism="Homo sapiens"
/db_xref="taxon:9606"
1.2481
/note="unnamed protein product"
/codon.start=1
/protein.id="CAD37798.1"
/db_xref="GI:21698762"
/translation="MEGAGGANDKKKISERRRKKSRDARSRRKSESEFYELAHQLPLPHNVSHLDKASVMRLTISYLRVRKLLDAGDLDEDMKQOMCFYALKDGFYMLTDDGMITYISDVNMYMGLTDFELTGHSGVDFPTCHDEHEMELTHNGYKKGKEONTORSEFLMKCTLSRGRTNFKSATKMYLCHGTHHIVYTNSSPOCGKKRPTCLVLCSEPIPHPSNIEPLDCKPFLSRHSIDMKFSCDERITFELMKPEPEELGKSTIYXHNALDSBHLTTHDMFTKGQVTTGGTIRMLAKRGIVWEQATVIYNTNSQPQCIVCVNTVSGITQHDILFSLQTECVLAKVESSDMKMTQLFTKVESEDTSLEDKLKEPALTLARAAADTILSIDFGSNDTETDQOLEEVLVNDVMSPNKLGNTIAMSPLETAETPRPLRSSADPALNOEVALKLEKLEAEDEAKNPSSTODTDLEMLAPYISPEPNSEPCFYVDSDVMEKLELVKLEAEDEAKNPSSTODTDLEMLAPYIPMDDFOLRSFDOLSPLESSASPEASPOSIVYVYVQOIOPEPTANATTTATDELKPYTKTKRMDIKILIASPPTNHHKTTTSPTPTDTSRATSPRKGVETETKSHRSRNVLSVALISQRTVYEEELNKLITALLNOAKRRMEHDGSLFAVGIQTLLQPDHAAFTLSMKRVKRGCKSSRQNGMEQTTILIPSDILACRLIGOSMDESGIPOLTSYDCRVNAPIGSRNMLDGEELRLALDVN"
BASE COUNT 829 a 512 c 500 g 640 t
ORIGIN
Query Match 93.6%; Score 23.4; DB 6; Length 2481;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 CCATGTGACCATGAGGAATGAGAG 25
Db 412 CCATGTGACCATGAGGAATGAGAG 436
RESULT 4
LOCUS AX481424 2481 bp DNA linear PAT 16-AUG-2002
DEFINITION Sequence 38 from Patent WO02055693.
ACCESSION AX481424

VERSION AX481424.1 GI:22316338
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
AUTHORS Kreutzer, R., Limmer, S., Rost, S. and Hadwiger, P.
TITLE Method for inhibiting the expression of a target gene
JOURNAL Patient: MO 02055693-A 38 18-JUL-2002;
Ribopharma AG (DE)
FEATURES
Source Location/Qualifiers
1..2481
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 829 a 512 c 500 g 640 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 6; Length 2481;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATGGAATGAGAG 25
|||||
Db 412 CCATGTGACCATGGAATGAGAG 436

RESULT 5
AF304431 2509 bp mRNA linear PRI 29-DEC-2000
LOCUS Homo sapiens hypoxia-inducible factor 1 alpha subunit (HIF1A) mRNA,
DEFINITION complete cds.
AF304431
VERSION AF304431.1 GI:11995454
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 2509)
AUTHORS Sun, B., Zhao, H.R., Yu, R.T. and Ni, M.S.H.
TITLE Direct Submision
JOURNAL Submitted (11-SEP-2000) Department of Neurosurgery, Affiliated Hospital of Xuzhou Medical College, Huaihai West Road, Xuzhou, Jiangsu 221002, China
FEATURES
Source Location/Qualifiers
1..2509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q21-q24"
/tissue_type="glioma"
1..2509
/gene="HIF1A"
1..2481
/gene="HIF1A"
/codon_start=1
/product="hypoxia-inducible factor 1 alpha subunit"
/protein_id="AA643026.1"
/db_xref="GI:11995454"
/translation="MEGAGGANDKKKISSRRKKESRDARSRRKSESEFYELAHOL
PLPHVSSHLDKASVMRLTISYLVRKLLDAGDIEDDKAKMNFYKALDGFYV
LTDGDMYISDNVNYKGLTQFELTGHVSFDTTHPDHDEMEMLTHRGVKKKE
ONTORSEFLRMKCTLTSGRTMNKSAWVFLCTGHIHYDTNSNQPGYKKPMT
CLVLCEPIPRPSNIEIPLDSKIFLSHSIDMKFSCYDEDEITELMGVEBELGRSIT
EYHALDSDLTKTHHDMFTKGOVTTQYRLAKRGYVAVETQATVAVINTKRSQOC
IVCNVYSGIIOHDLFSIQTECVLKPVESSDMKWTOLFTVSESDSLEFDKAK
EPDALTLLAPAGDTIISLDGSDNDETDQOLEVPLINDVLPSPNKLONINLAK
SPLETAETPKPLRSSADPALNOEVALKELEPELSEFTMPQIODOTSPSGSTRQ
SSPSPNSSEYCFVDSMDVNEKLELVEKLEFADTEAKNPSTOTDIDLEMLAYI
PMDDPOLRSFDOLSPLESSASPEASPOSITVVOQIOEPTANATTATTTATDEL
KTVKRDMEIKILIASPETHHKTSTSSPYDTSRTASPNRAGGVIEQTEK
SHRSPVLSVALSQRITVEBELNPKITLALQNAOKRMEHDSILFOAVGIGTLLOO
DCEVNAPIQSSRNILQEEELRALDOVN"

BASE COUNT 832 a 517 c 502 g 658 t
ORIGIN
PDDHAATTSLSMKRVKGGKSSQNGMECKTITILPSDLACRLLGSMDESGLPOLTSY
DCEVNAPIQSSRNILQEEELRALDOVN"

Query Match 93.6%; Score 23.4; DB 9; Length 2509;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCATGTGACCATGGAATGAGAG 25
|||||
Db 412 CCATGTGACCATGGAATGAGAG 436

RESULT 6
AF208487 2522 bp DNA linear PRI 27-DEC-1999
LOCUS Homo sapiens hypoxia-inducible factor 1 alpha (HIF1A) gene,
DEFINITION complete cds.
AF208487
VERSION AF208487.1 GI:6636337
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 2522)
AUTHORS Rupert, J.L. and Hochachka, P.W.
TITLE HIF1A sequence in the Quechua, a high altitude population
JOURNAL Unpublished
2 (bases 1 to 2522)
Rupert, J.L. and Hochachka, P.W.
DIRECT SUBMISSION
Submitted (25-NOV-1999) Zoology, University of British Columbia,
6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
FEATURES
Source Location/Qualifiers
1..2522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q21-q24"
/cell_line="GM1198 (Cortell)"
/note="derived from a high altitude native (Quechua)"
11..2491
/gene="HIF1A"
/note="HIF1-alpha"
11..2491
/gene="HIF1A"
/note="transcription factor: hypoxia-inducible factor 1a"
/codon_start=1
/product="hypoxia-inducible factor 1 alpha"
/protein_id="AAF20149.1"
/db_xref="GI:6636338"
/translation="MEGAGGANDKKKISSRRKKESRDARSRRKSESEFYELAHOL
PLPHVSSHLDKASVMRLTISYLVRKLLDAGDIEDDKAKMNFYKALDGFYV
LTDGDMYISDNVNYKGLTQFELTGHVSFDTTHPDHDEMEMLTHRGVKKKE
ONTORSEFLRMKCTLTSGRTMNKSAWVFLCTGHIHYDTNSNQPGYKKPMT
CLVLCEPIPRPSNIEIPLDSKIFLSHSIDMKFSCYDEDEITELMGVEBELGRSIT
EYHALDSDLTKTHHDMFTKGOVTTQYRLAKRGYVAVETQATVAVINTKRSQOC
IVCNVYSGIIOHDLFSIQTECVLKPVESSDMKWTOLFTVSESDSLEFDKAK
EPDALTLLAPAGDTIISLDGSDNDETDQOLEVPLINDVLPSPNKLONINLAK
SPLETAETPKPLRSSADPALNOEVALKELEPELSEFTMPQIODOTSPSGSTRQ
SSPSPNSSEYCFVDSMDVNEKLELVEKLEFADTEAKNPSTOTDIDLEMLAYI
PMDDPOLRSFDOLSPLESSASPEASPOSITVVOQIOEPTANATTATTTATDEL
KTVKRDMEIKILIASPETHHKTSTSSPYDTSRTASPNRAGGVIEQTEK
SHRSPVLSVALSQRITVEBELNPKITLALQNAOKRMEHDSILFOAVGIGTLLOO
DCEVNAPIQSSRNILQEEELRALDOVN"

Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCATGTGACCATTTAGGAATGAGAG 25
|||||
Db 422 CCATGTGACCATTTAGGAATGAGAG 446

RESULT 7
AX230580 2528 bp DNA PAT 11-SEP-2001
LOCUS Sequence 2 from Patent W00162965.
ACCESSION AX230580
VERSION AX230580.1 GI:15592425
KEYWORDS human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2528)
Kingsman, A.J.
TITLE Differential expression screening method
JOURNAL Patent: WO 0162965-A 2 30-AUG-2001;
Oxford Biomedica (UK) Limited (GB)
FEATURES
source Location/Qualifiers
1..2528
/db_xref="taxon:9606"
BASE COUNT 839 a 528 c 513 g 648 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 6; Length 2528;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCATGTGACCATTTAGGAATGAGAG 25
|||||
Db 439 CCATGTGACCATTTAGGAATGAGAG 463

RESULT 8
AF207601 2530 bp mRNA linear PRI 27-DEC-1999
LOCUS Homo sapiens cell-line GM1201 hypoxia-inducible factor 1 alpha
DEFINITION (HIF1A) mRNA, complete cds.
ACCESSION AF207601
VERSION AF207601.1 GI:6636316
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2530)
Rupert, J.L. and Hochachka, P.W.
TITLE HIF1a sequence in the Quechua, a high altitude population
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2530)
Rupert, J.L. and Hochachka, P.W.
REPRINTS Direct Submission
TITLE Submitted (18-NOV-1999) Zoology, University of British Columbia,
JOURNAL 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
FEATURES
source Location/Qualifiers
1..2530
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q21-q24"
/cell_line="GM1201 (Coriell)"
/cell_type="lymphoblast"
/note="derived from a high altitude native (Quechua)"
1..2530
/gene="HIF1A"
/note="HIF1-alpha"

CDS

11..2491
/gene="HIF1A"
/note="transcription factor; hypoxia-inducible factor 1a"
/codon_start=1
/product="hypoxia-inducible factor 1 alpha"
/protein_id="AA20139.1"
/db_xref="GI:6636317"
/translation="MEGAGAGANDKRRKISSERRKSRDARSRSKSEVYELAHOL
PLPHNYSHLDRASVRLTISYLRYKRLKLDGDDLEDDMKAGNVCYLKALDSFVW
LTPDGMVTSIDNVNKMVGLTQELTGHVSFEDTPRCDHEMRRLTHRNGLVKGKE
QNTQRFPLRMKCTLSRGTRTNMKSATMKVHCHGTHVDTNSNPOCCKKPMPT
CLVLTCEPDPHPSNIEIPLDSTFTLSRSIDMKFSYCDERTELMGYPPELLGRSY
EYVHALDSHLTKTHHMTKQVTTGGOTRMLARKGVYVWETQATVYNTKNSPOPC
IVCVNVVSGIIQHDLIIFSLQTECVLKEVESDDMKMTQLTKVESSEPTSLPKIK
EPDALTILAPAGDITISIDFGSNDTETDDQLEVPYINVMPSFVKELQINLMA
SPLETPETPKPLRSSADPALNOEVALKEPNESELESTMPQIDQTPSPDSSTRO
SPEPNSPSPCYVDSDVWNEKLEVLKLEPAEDTEAKNPFSTQDLDLEMLAPYI
PMDDPOLRSEFDLSPLSSSSASPEASPSQVYVVOOQIOEPANATTTATYDEL
KVTQKDMEDIKLILASPTTHHKETTSATSPYRDOQSRTASPNRAGKGVIRQTEK
SHRSPNVLSVALSORVYVEELNKILALONAKRKMEDSLPQVIGITLQO
PDDHAATTSLSMKRVKCKSEQNGEOKTITILPSDLACRLGQSDPSGLPOLTSY
DCEVNAPIQGRNMLQGEELLRALDOVN"

BASE COUNT 836 a 524 c 508 g 662 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 9; Length 2530;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCATGTGACCATTTAGGAATGAGAG 25
|||||
Db 422 CCATGTGACCATTTAGGAATGAGAG 446

RESULT 9
AF207602 2537 bp mRNA linear PRI 27-DEC-1999
LOCUS Homo sapiens cell-line GM1201 hypoxia-inducible factor 1 alpha
DEFINITION (HIF1A) mRNA, complete cds.
ACCESSION AF207602
VERSION AF207602.1 GI:6636318
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2537)
Rupert, J.L. and Hochachka, P.W.
TITLE HIF1a sequence in the Quechua, a high altitude population
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2537)
Rupert, J.L. and Hochachka, P.W.
REPRINTS Direct Submission
TITLE Submitted (18-NOV-1999) Zoology, University of British Columbia,
JOURNAL 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
FEATURES
source Location/Qualifiers
1..2537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q21-q24"
/cell_line="GM1197 (Coriell)"
/cell_type="lymphoblast"
/note="derived from a high altitude native (Quechua)"
1..2537
/gene="HIF1A"
/note="HIF1-alpha"

CDS

8..2488
/gene="HIF1A"
/note="transcription factor; hypoxia-inducible factor 1a"
/codon_start=1
/product="hypoxia-inducible factor 1 alpha"
/protein_id="AA20140.1"

```

/db_xref="GI:6636319"
/translation="MEGAGAGANDKKKISSERRRKKESRDAARRSRKESVFEYELAHOL
PLPHNVSSHLDKASVMRLTISYLVRKRLDAGDLDIEDMAKQMCFTYKALDGVNV
LTDGDMITISDVNKKYGLTQFELTGHVSFDFTHPCDHEERMLTHRNGLVKRGKE
QNTQSFELPRKCTILSRGRNTNKSATWKLHCTGHIHYVDTNSNPQCGYKPPMT
CVLICEPIPHSNIEIPLDSTKFLSRHSLDMKFSYCDERITELMGYEPELGRSII
EYHALDSHDLTKTHHDMFTKGQVTRGYRLARKGYVWVETQATVINYTKDSQPC
IVCNVVVGSGIIQHDILFSLQOTECVLRKPVESDMKMTOLFTKVESDTSCLPKLK
EPDALLAPAGDITISIDGSDTETDQLEVPYINDVMSPNKCLONILAM
SELPATETPKLRSSADPALNDFEVLKLEAEDEAKNPESADTDLDEMLAPYIP
SSPEPSSEYCFYVDSMDVNEFKLELAEDEAKNPESADTDLDEMLAPYIP
PMDDFQLRSEFDLSPLESSASPESSASTVYVQIQOIEPTANATTAATTTDEL
KVTQKDMEDIKILIASPSPTHHKETTSPTSPYEDQSPRSPNRAKGYIETDTEK
SHRSPNVALSISORTYVEEELNPTIALONAKRKMEHDSGLFOAGIGTILQO
PDHATTSLSWKRKVGKSSQNGNEQKTIILIFSDLACRLGQSMDSGLPQLTST
DCEVNAPIQGSNRLDGEELRALDOVN"
BASE COUNT      838 a      526 c      507 g      666 t
ORIGIN
Query Match      93.6%; Score 23.4; DB 9; Length 2537;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCATGTGACCATTAGAAGATGAG 25
Db 419 CCATGTGACCATGAGGAATGAG 443
RESULT 10
LOCUS      RNHPFAC1      2711 bp      mRNA      linear      ROD 26-FEB-2001
DEFINITION      R.norvegicus mRNA for hypoxia-inducible factor 1.
VERSION      Y09507
KEYWORDS      hypoxia-inducible factor 1.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
REFERENCE
AUTHORS      Kietzmann,T., Cornesse,Y., Brechtel,K., Modaresi,S. and
Jungermann,K.
TITLE      Perivenous expression of the mRNA of the three hypoxia-inducible
factor alpha-subunits, HIF1alpha, HIF2alpha and HIF3alpha, in rat
liver
JOURNAL      Biochem. J. 354 (Pt 3), 531-537 (2001)
MEDLINE      21134367
PUBMED      11237857
REFERENCE
AUTHORS      Kietzmann,T.
TITLE      Direct Submission
JOURNAL      Submitted (18-NOV-1996) T. Kietzmann, Inst. Of Biochem. And Mol.
Cell Biol., Georg-August-Univ. Goettingen, Humboldtallee 23,
D-37073 Goettingen, FRG
REMARK
AUTHORS      Kietzmann,T.
TITLE      Revised by [3]
JOURNAL      3 (bases 1 to 2711)
REFERENCE
AUTHORS      Direct Submission
TITLE      Submitted (27-FEB-1998) T. Kietzmann, Inst. Of Biochem. And Mol.
Cell Biol., Georg-August-Univ. Goettingen, Humboldtallee 23,
D-37073 Goettingen, FRG
COMMENT
FEATURES
SOURCE      location/Qualifiers
1..2711
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/sex="male"
/cell_line="primary hepatocytes"
/dev_stage="adult"
124..2601
/codon_start=1

```

```

/product="hypoxia-inducible factor 1"
/protein_id="CAA0701.1"
/db_xref="GI:2632131"
/db_xref="SPTREMBL:035800"
/translation="MEGAGCENKKKKSSERRRKKESRDAARRSRKESVFEYELAHOL
PLPHNVSSHLDKASVMRLTISYLVRKRLDAGDLDIEDMAKQMCFTYKALDGVNV
LTDGDMITISDVNKKYGLTQFELTGHVSFDFTHPCDHEERMLTHRNGLVKRGKE
QNTQSFELPRKCTILSRGRNTNKSATWKLHCTGHIHYVDTNSNPQCGYKPPMT
CVLICEPIPHSNIEIPLDSTKFLSRHSLDMKFSYCDERITELMGYEPELGRSII
EYHALDSHDLTKTHHDMFTKGQVTRGYRLARKGYVWVETQATVINYTKDSQPC
IVCNVVVGSGIIQHDILFSLQOTECVLRKPVESDMKMTOLFTKVESDTSCLPKLK
EPDALLAPAGDITISIDGSDTETDQLEVPYINDVMSPNKCLONILAM
SELPATETPKLRSSADPALNDFEVLKLEAEDEAKNPESADTDLDEMLAPYIP
SSPEPSSEYCFYVDSMDVNEFKLELAEDEAKNPESADTDLDEMLAPYIP
PMDDFQLRSEFDLSPLESSASPESSASTVYVQIQOIEPTANATTAATTTDES
KAVTKDMEDIKILIASPSPTHHKETTSPTSPYEDQSPRSPNRAKGYIETDTEK
SHRSPNVALSISORTYVEEELNPTIALONAKRKMEHDSGLFOAGIGTILQO
PDHATTSLSWKRKVGKSSQNGNEQKTIILIFSDLACRLGQSMDSGLPQLTST
DCEVNAPIQGSNRLDGEELRALDOVN"
BASE COUNT      798 a      613 c      623 g      677 t
ORIGIN
Query Match      93.6%; Score 23.4; DB 10; Length 2711;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCATGTGACCATTAGAAGATGAG 25
Db 535 CCATGTGACCATGAGGAATGAG 559
RESULT 11
LOCUS      AB073325      3551 bp      mRNA      linear      PRI 23-OCT-2001
DEFINITION      Homo sapiens HIF1A mRNA for hypoxia-inducible factor 1 alpha
variant, complete cds.
VERSION      AB073325
KEYWORDS      AB073325.1 GI:16326343
SOURCE      Homo sapiens liver cDNA to mRNA.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Tanaka,S. and Sugimachi,K.
TITLE      Hypoxia-inducible factor-1 alpha variant isolated from human liver
tissue
JOURNAL      unpublished
REFERENCE
AUTHORS      Tanaka,S. and Sugimachi,K.
TITLE      Direct Submission
JOURNAL      Submitted (20-OCT-2001) Shinji Tanaka, Kyushu University, Graduate
School of Medical Sciences, Department of Surgery and Science;
3-1-1 Maidashi, Fukuoka 812-8582, Japan
(E-mail:shinji@surg2.med.kyushu-u.ac.jp, Tel:81-92-642-5466,
Fax:81-92-642-5482)
FEATURES
SOURCE      Location/Qualifiers
1..3551
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
1..3551
/gene="HIF1A"
29..2236
/gene="HIF1A"
/note="HIF-1 alpha variant"
/codon_start=1
/product="hypoxia-inducible factor 1 alpha variant"
/protein_id="BAB70608.1"
/db_xref="GI:16326344"
/translation="MEGAGCAGANDKKKISSERRRKKESRDAARRSRKESVFEYELAHOL
PLPHNVSSHLDKASVMRLTISYLVRKRLDAGDLDIEDMAKQMCFTYKALDGVNV
LTDGDMITISDVNKKYGLTQFELTGHVSFDFTHPCDHEERMLTHRNGLVKRGKE

```

ONTORSEFELMKKCTLTSGRGTMTNKATFWVLHCTGHIHYVDINSNOPOGKYKPPMT
CIVILCEPIHPSPNIEIPLDSKTEFLSHSDMKESYCDERITELMGEPPELLGRSIV
EYHALDSDLTHTHDMFTKQVTTQYMLAKRGYVWEVQATYIYNTKRSQOC
IVCNVYVSGIIOHDLFTSLQOTECVLRKPESSDMKTQLTVESDITSLDKLK
EPDALTLAPAGDTIISLDGSDNETEDDOQLEAPLYDMVLPSPNELOMINIAM
SPLEATETPKPRSSADPALNOEVALKLEPNPSPLESTFMPDIODOTSPSGSTRO
SSPEPNSPEYCEFDVSDMNEFKLEVEKLFADTEPAKPFSTODTDLDEMLAYI
PMDDPOLRSFSDSPSSASPEASPOSTVYVROQOIOEPTANATTTATTEL
KYTKRDMEDIKILIASPPPHIHKETTSSTSPYDTSRTASPNRAGGYEOTEK
SHPSRPNVLSVALSQRTPPEBELPRTLLQNAQRKKRMEHDSLFQAVGII"

BASE COUNT 1150 a 671 c 650 g 1080 t

Query Match 93.6%; Score 23.4; DB 9; Length 3551;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CCATGTGACCATTAAGAAATGAGAG 25
|||||
Db 440 CCATGTGACCATTAAGAAATGAGAG 464

RESULT 12
LOCUS HSU22431
DEFINITION Human hypoxia-inducible factor 1 alpha (HIF-1 alpha) mRNA, complete cds.
ACCESSION U22431
VERSION U22431.1 GI:881345
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3678)
Wang, G.L., Jiang, B.-H., Rue, E.A. and Semenza, G.L.
Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension
Proc. Natl. Acad. Sci. U.S.A. 92 (12), 5510-5514 (1995)
JOURNAL MEDLINE 95296340
PUBMED 7539918
REFERENCE 2 (bases 1 to 3678)
Wang, G.L., Jiang, B.-H., Rue, E.A. and Semenza, G.L.
Direct Submission
Submitted (09-MAR-1995) Gregg L. Semenza, Center for Medical Genetics, The Johns Hopkins University School of Medicine, 600 N. Wolfe St., Baltimore, MD 21287-3914, USA
JOURNAL

FEATURES
source
1. 3678
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Hep3B"
/cell_type="hepatoblastoma"
1. 3678
/gene="HIF-1 alpha"
29. 2509
/gene="HIF-1 alpha"
/standard_name="hypoxia-inducible factor 1, alpha subunit"
/note="basic helix-loop-helix transcription factor."
/codon_start=1
/product="hypoxia-inducible factor 1 alpha"
/protein_id="AAC50152.1"
/db_xref="GI:881346"
/translation="MEGAGAGANDKKKISSRRKESRDAAARRSRKSESEVFEYLAHOL
PLPHNSVSHLDKASVRLTISYLRKILGAGDIDEDDKAQNCFYLALDGFVAV
LTDDGDMYIISDNVKKYGLTGFELTGHVDFTHPDHEMRMLTHRGVAKKKE
ONTORSEFELMKKCTLTSGRGTMTNKATFWVLHCTGHIHYVDINSNOPOGKYKPPMT
CIVILCEPIHPSPNIEIPLDSKTEFLSHSDMKESYCDERITELMGEPPELLGRSIV
EYHALDSDLTHTHDMFTKQVTTQYMLAKRGYVWEVQATYIYNTKRSQOC
IVCNVYVSGIIOHDLFTSLQOTECVLRKPESSDMKTQLTVESDITSLDKLK
EPDALTLAPAGDTIISLDGSDNETEDDOQLEAPLYDMVLPSPNELOMINIAM
SPLEATETPKPRSSADPALNOEVALKLEPNPSPLESTFMPDIODOTSPSGSTRO
SSPEPNSPEYCEFDVSDMNEFKLEVEKLFADTEPAKPFSTODTDLDEMLAYI

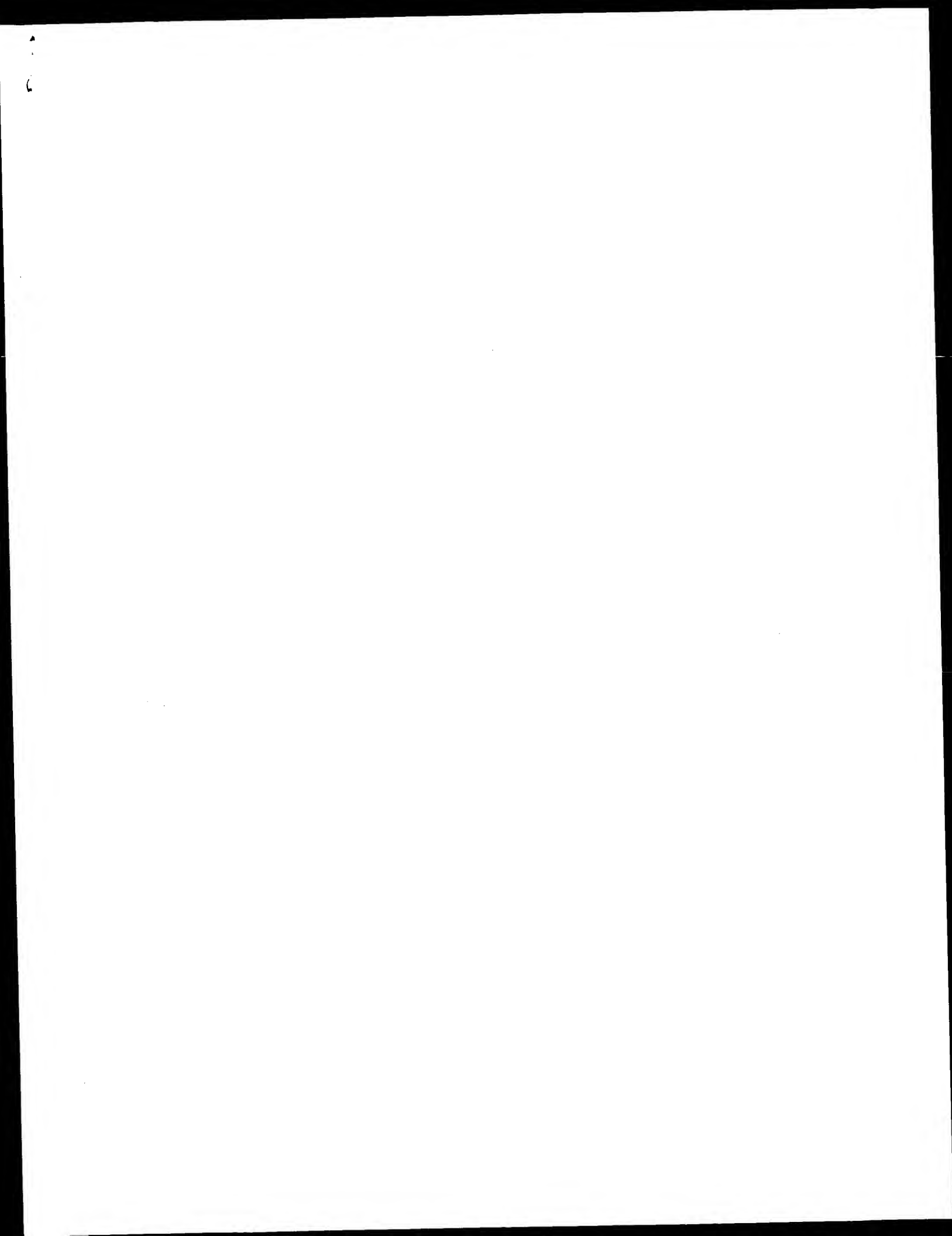
polyA_site
/gene="HIF-1 alpha"
/note="42 A nucleotides"
BASE COUNT 1197 a 695 c 675 g 1111 t

Query Match 93.6%; Score 23.4; DB 9; Length 3678;
Best Local Similarity 96.0%; Pred. No. 0.38;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CCATGTGACCATTAAGAAATGAGAG 25
|||||
Db 440 CCATGTGACCATTAAGAAATGAGAG 464

RESULT 13
LOCUS AF057308
DEFINITION Rattus norvegicus hypoxia-inducible factor-1 alpha (Hif1a) mRNA, complete cds.
ACCESSION AF057308
VERSION AF057308.1 GI:4580532
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3718)
Zou, A.P., Yang, Z.Z., Li, P.L. and Cowley, A.W. Jr.
Oxygen-dependent expression of hypoxia-inducible factor-1alpha in renal medullary cells of rats
Physiol. Genomics (Online) 6 (3), 159-168 (2001)
JOURNAL MEDLINE 21417706
PUBMED 11526200
REFERENCE 2 (bases 1 to 3718)
Zou, A.P., Su, N., Park, F., Li, P.L. and Cowley, A.W. Jr.
Direct Submission
Submitted (04-APR-1998) Departments of Physiology and Pharmacology, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA
JOURNAL

FEATURES
source
1. 3718
/organism="Rattus norvegicus"
/strain="Sprague Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="Kidney"
/dev_stage="adult"
1. 3718
/gene="Hif1a"
24. 2495
/gene="Hif1a"
/note="transcription factor; RHIF-1"
/codon_start=1
/product="hypoxia-inducible factor-1 alpha"
/protein_id="AAD24413.1"
/db_xref="GI:4580533"
/translation="MEGAGAGENEKKNRMSERRKESRDAAARRSRKSESEVFEYLAHOL
PLPHNSVSHLDKASVRLTISYLRKILGAGDIDEDDKAQNCFYLALDGFVAV
LTDDGDMYIISDNVKKYGLTGFELTGHVDFTHPDHEMRMLTHRGVAKKKE
EONTORSEFELMKKCTLTSGRGTMTNKATFWVLHCTGHIHYVDINSNOPOGKYKPPMT
CIVILCEPIHPSPNIEIPLDSKTEFLSHSDMKESYCDERITELMGEPPELLGRSIV
EYHALDSDLTHTHDMFTKQVTTQYMLAKRGYVWEVQATYIYNTKRSQOC
IVCNVYVSGIIOHDLFTSLQOTECVLRKPESSDMKTQLTVESDITSLDKLK
EPDALTLAPAGDTIISLDGSDNETEDDOQLEAPLYDMVLPSPNELOMINIAM
SPLEATETPKPRSSADPALNOEVALKLEPNPSPLESTFMPDIODOTSPSGSTRO
SSPEPNSPEYCEFDVSDMNEFKLEVEKLFADTEPAKPFSTODTDLDEMLAYI



[illegible]

XX (ASPE-) ASPERA PHARM AB.
 PA
 XX
 XX Poellinger L, Pereira T, Ruas J;
 PI
 XX WPI: 2002-257466/30.
 DR
 DR P-PSDB: AA077602.
 XX
 PT New polypeptides comprising hypoxia-inducible factor-1 with alterations
 PT of the transactivation domain, useful treating ischemic conditions,
 PT e.g. brain infarction, heart infarction or circulatory disorder -
 PS
 PS Disclosure: Page 73-76; 80pp; English.
 XX
 XX The invention relates to a polypeptide comprising hypoxia-inducible
 CC factor-1 (HIF-1) with alterations of the transactivation domain (N-TAD or
 CC C-TAD). Also included are nucleic acids encoding the altered proteins, a
 CC vector comprising the nucleic acid, a host cell transformed with the
 CC vector, methods for producing the protein or its functional fragment
 CC or an isolated degradation box, a method of screening for an agent that
 CC modulates N-TAD function and antagonists, agonists, modulators and
 CC HIF-1 peptide fragments useful for modulating HIF-1 function or the
 CC function of proteins that interact with it. The isolated polypeptides and
 CC their fragments with altered residues are useful in methods for treating
 CC diseases. The disease is an ischemic condition, e.g. brain infarction,
 CC heart infarction or circulatory disorder. The disease may also be cancer,
 CC hypertension, demyelinating disorders, diffuse proliferative
 CC glomerulonephritis, toxoplasmosis caused retinochoroiditis, HIV (human
 CC immunodeficiency virus) caused T4 angioinosis, HIV-caused Kaposi's
 CC sarcoma, hepatitis-caused inflammation, hepatitis-caused angioinosis,
 CC chronic ulceration, proliferative retinopathy, retina haemangioblastomas,
 CC neovascularisation, arterial hypervascularisation, sarcoidosis, bullous
 CC skin disease, vasculitis with angioinosis, dermatomyositis with
 CC angioinosis, polymyositis with angioinosis, rheumatoid arthritis,
 CC juvenile osteoarthritis, polyarthritits, aneurysm or atheroma. The
 CC present sequence encodes HIF-1.
 CC
 SQ Sequence 2481 BP; 829 A; 512 C; 500 G; 640 T; 0 other;
 QY
 Query Match 93.6%; Score 23.4; DB 24; Length 2481;
 Best Local Similarity 96.0%; Pred. No. 0.19;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 DB 412 CCATGTGACCATTTAGGAATGAGAG 25
 1 CCATGTGACCATTTAGGAATGAGAG 25
 ||||||||||| |||||||||||
 412 CCATGTGACCATTTAGGAATGAGAG 436
 RESULT 2
 ABL91695
 ID ABL91695 standard; DNA: 2481 BP.
 AC
 AC ABL91695;
 XX
 XX 28-MAY-2002 (first entry)
 DT
 XX
 XX Human polynucleotide SEQ ID NO 38.
 DE
 XX
 XX Human; HIV; HCV; gene expression; oligoribonucleotide; tumour; pathogen;
 KW Plasmidium; virus; viroid; cytokine; prion; antisense oligonucleotide;
 KW cytosolic; virucide; protozoacide; antibacterial; ds.
 XX
 OS Homo sapiens.
 XX
 XX DE10100586-C1.
 FN
 FN 11-APR-2002.
 PD
 PD 09-JAN-2001; 2001DE-1000586.
 PF
 PF 09-JAN-2001; 2001DE-1000586.
 PR
 PR (RIBO-) RIBOPHARMA AG.
 PA

XX
 PI Kreutzer R, Limmer S, Rost S, Hadwiger P;
 XX
 XX WPI: 2002-270454/32.
 DR
 XX
 XX Inhibiting gene expression in cells, useful for e.g. treating tumors,
 PT by introducing double-stranded complementary oligorRNA having unpaired
 PT terminal bases -
 PS
 PS Claim 13; Page 33; 104pp; German.
 XX
 XX The invention relates to a method for inhibiting expression of a target
 CC gene (ABL91658-ABL91797) in a cell by introducing at least one
 CC oligoribonucleotide that has a double-stranded structure consisting of at
 CC most 49 sequential nucleotide pairs, with at least part of one strand
 CC complementary with the target gene and has at least one end a
 CC single-stranded segment of 1-4 nt. The method provides
 CC oligoribonucleotides for antisense inhibition of gene expression useful
 CC e.g. for treating tumors but the oligoribonucleotides may also be
 CC directed against genes present in pathogens (e.g. Plasmidium or
 CC viruses/viroids, pathogenic on humans, animals or plants) or against
 CC cytokine, Id, developmental or prion genes. The method provides more
 CC effective inhibition of gene-expression than use of known
 CC oligonucleotides, probably because the unpaired overhang increases
 CC stability and thus intracellular concentration.
 CC
 SQ Sequence 2481 BP; 829 A; 512 C; 500 G; 640 T; 0 other;
 QY
 Query Match 93.6%; Score 23.4; DB 24; Length 2481;
 Best Local Similarity 96.0%; Pred. No. 0.19;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 DB 412 CCATGTGACCATTTAGGAATGAGAG 25
 1 CCATGTGACCATTTAGGAATGAGAG 25
 ||||||||||| |||||||||||
 412 CCATGTGACCATTTAGGAATGAGAG 436

RESULT 3
 AAS14154
 ID AAS14154 standard; CDNA: 2528 BP.
 XX
 AC AAS14154;
 XX
 XX 18-DEC-2001 (first entry)
 DT
 XX
 XX Human HIF-1 alpha DNA used in identification of hypoxia regulated genes.
 DE
 XX
 XX Differential expression; polymorphism; biological pathway; IRES; GPP; ss;
 KW internal ribosome entry site; green fluorescent protein; HIF-1 alpha;
 KW hypoxia inducible factor 1 alpha; hypoxia regulated gene; macrophage;
 KW human.
 KW
 OS Homo sapiens.
 XX
 XX WO200162965-A2.
 PN
 PN 30-AUG-2001.
 PD
 XX
 XX 22-FEB-2001; 2001WO-GB00758.
 PF
 PF 22-FEB-2000; 2000GB-0004197.
 PR
 PR 28-JUL-2000; 2000GB-0018679.
 XX
 XX (OXFO-) OXFORD BIOMEDICA UK LTD.
 PA
 PA Kingsman AJ;
 PI
 PI WPI: 2001-589807/66.
 DR
 XX
 XX Screening a genetic element involved in a cellular process, comprises
 PT comparing gene expressions in a cell, and a second cell that has
 PT altered levels of genes used in the process, and detecting an element
 PT with varied expression -

CC binding molecules or HIF-1-SUMO-1 complex modulators. mdrl-hypoxia
 CC responsive element (HRE) binding molecules or antisense nucleic
 CC acid molecules and SUMO-1 binding molecules or antisense molecules
 CC are useful for treating a subject having or at risk of developing
 CC haematologic malignancy or MDR (e.g. a lymphoid disorder or a myeloid
 CC disorder). The lymphoid disorders include lymphocytic leukaemia or
 CC chronic lymphoproliferative disorders e.g. lymphoma, myeloma or chronic
 CC lymphoid leukaemia. The myeloid disorders include chronic or acute
 CC myeloid leukaemia, e.g. angiotenic myeloid metaplasia, essential
 CC thrombocythaemia or polycythaemia vera. The invention is used in gene
 CC therapy. The present sequence is human HIF-1alpha cDNA.
 CC
 SQ Sequence 3678 BP; 1197 A; 695 C; 675 G; 1111 T; 0 other;
 Query Match 93.6%; Score 23.4; DB 24; Length 3678;
 Best Local Similarity 96.0%; Pred. No. 0.2;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CCATGTGACCATAGGAATGAGAG 25
 DB 440 CCATGTGACCATAGGAATGAGAG 464
 RESULT 6
 ABR84267
 ID ABR84267 standard; cDNA: 3678 BP.
 AC ABR84267;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #838.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 KW
 XX
 OS Homo sapiens.
 OS
 XX
 PN WO200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US30821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX
 DR WPI; 2002-435328/46.
 XX
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX
 PS Claim 1; SEQ ID NO 838; 114pp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GCA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially

CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GCA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 3678 BP; 1197 A; 695 C; 675 G; 1111 T; 0 other;
 Query Match 93.6%; Score 23.4; DB 24; Length 3678;
 Best Local Similarity 96.0%; Pred. No. 0.2;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CCATGTGACCATAGGAATGAGAG 25
 DB 440 CCATGTGACCATAGGAATGAGAG 464
 RESULT 7
 AAT45937
 ID AAT45937 standard; DNA: 3736 BP.
 AC AAT45937;
 XX
 DT 19-MAR-1997 (first entry)
 XX
 DE Human hypoxia inducible factor-1 alpha cDNA.
 XX
 KW Hypoxia inducible factor-1 alpha; HIF-1; tissue damage;
 KW atherosclerosis; cerebral artery disease; gene therapy; ss.
 KW
 XX
 OS Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FH
 FT msc_signal 23..32 /tag= a
 FT
 FT CDS 29..2509 //function= kozak sequence
 FT /tag= b
 FT polyA_signal 3674..3679 /tag= c
 FT
 XX
 PN WO9639426-A1.
 XX
 PD 12-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US10251.
 XX
 PR 06-JUN-1995; 95US-0480473.
 XX
 PA (UyGO) UNIV JOHNS HOPKINS SCHOOL MED.

XX PI Semenaza GL;
 XX DR MPI: 1997-043061/04.
 XX DR P-PSDB; AAM06557.
 PT DNA encoding human hypoxia-inducible factor 1 alpha - useful for
 PT enhancing expression of structural gene and treatment of
 PT hypoxia-related tissue damage
 XX
 PS Claim 2; Page 49-53; 95pp; English.
 CC A DNA sequence (AAT45937) codes for the alpha subunit (AAM06557) of
 CC human hypoxia inducible factor-1 (HIF-1), a DNA binding protein
 CC that binds to the enhancer region of e.g. erythropoietin and
 CC vascular endothelial growth factor (VEGF) genes. The DNA sequence
 CC is a composite of 3 clones obt. by screening an Hep3b library and
 CC by database analysis. HIF-1 alpha in transformed host cells, as
 CC prepare recombinant HIF-1 alpha in transfected host cells, as
 CC probes, and in the gene therapy of HIF-1-mediated or hypoxia-related
 CC disorders such as atherosclerotic coronary or cerebral artery
 CC disease; antisense sequences inhibit HIF-1 expression e.g. to treat
 CC tumour proliferation mediated by VEGF-induced angiogenesis.
 CC
 SQ Sequence 3736 BP; 1255 A; 695 C; 675 G; 1111 T; 0 other;
 Query Match 93.6%; Score 23.4; DB 18; Length 3736;
 Best Local Similarity 96.0%; Pred. No. 0.2;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 CCATGTGACCATTTAGGAATGAGAG 25
 DB 440 CCATGTGACCATTTAGGAATGAGAG 464
 RESULT 8
 AA29537
 ID AA29537 standard; DNA; 3736 BP.
 AC AA29537;
 XX
 DT 03-JUL-2000 (first entry)
 DE DNA encoding a wild type human hypoxia inducible factor-1 alpha.
 XX
 KW Human; hypoxia-inducible factor 1 alpha; HIF-1alpha; variant;
 KW hypoxia inducible gene; hypoxia inducible factor; hypoxia;
 KW ischemia related damage; angiogenesis; coronary artery disease;
 KW ischemic tissue damage; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 29..2509
 FT /*tag= a
 FT /product= "hypoxia inducible factor-1 alpha"
 XX
 PN WO200010578-A1.
 XX
 PD 02-MAR-2000.
 XX
 PF 25-AUG-1999; 99WO-US19416.
 XX
 PR 25-AUG-1998; 98US-0148547.
 XX
 PA (UW30) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 PI Semenaza GL;
 XX MPI: 2000-246493/21.
 XX DR P-PSDB; AAY69407.
 XX
 PT Variant forms of hypoxia-inducible factor (HIF)-1 alpha, useful for

PT treating hypoxia or ischemia-related tissue damage -
 XX
 PS Disclosure; Page 80-89; 96pp; English.
 CC
 CC The present sequence encodes a wild type human hypoxia-inducible factor
 CC (HIF)-1 alpha. The specification describes HIF-1alpha variants that are
 CC stable under hypoxic and non-hypoxic conditions. The variants comprises
 CC amino acid residues 1-391 and 521-826, 549-826, 576-826, 429-826,
 CC 469-826, 494-826, 508-826, 512-826 or 517-826 of the wild type human
 CC HIF-1alpha polypeptide, in which residues 551 and 552 are not setine
 CC and threonine, respectively. The HIF-1alpha variant polynucleotide
 CC sequences are useful for increasing expression of a hypoxia inducible
 CC gene in a cell. They is also useful for providing constitutive
 CC or preventing hypoxia or ischemia related damage. The variant
 CC HIF-1alpha polypeptides are useful for providing prophylactic therapy
 CC for inducing the level of angiogenesis in tissues of patients at risk
 CC of coronary artery disease or ischemic tissue damage.
 CC
 SQ Sequence 3736 BP; 1255 A; 695 C; 675 G; 1111 T; 0 other;
 Query Match 93.6%; Score 23.4; DB 21; Length 3736;
 Best Local Similarity 96.0%; Pred. No. 0.2;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 CCATGTGACCATTTAGGAATGAGAG 25
 DB 440 CCATGTGACCATTTAGGAATGAGAG 464
 RESULT 9
 AB199710
 ID AB199710 standard; CDNA; 3746 BP.
 AC AB199710;
 XX
 DT 07-MAR-2002 (first entry)
 DE Mouse ischaemic condition related CDNA sequence SEQ ID NO:759.
 XX
 KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
 XX
 OS Mus musculus.
 XX
 PN WO2001088188-A2.
 XX
 PD 22-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-JF04192.
 XX
 PR 18-MAY-2000; 2000JP-0145977.
 XX
 PA (UWNT-) UNIV NIHON SCHOOL JURIDICAL PERSON.
 PI Ishikawa K, Asai S, Takehashi Y, Nagata T, Ishii Y;
 XX
 DR MPI: 2002-034733/04.
 XX
 DR P-PSDB; ABB57270.
 XX
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or
 PT by determining the expression profile of a gene group comprising these
 PT genes -
 XX
 PS Claim 2; Page 1863-1869; 2690pp; English.
 CC
 CC The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (I) in a test sample or determining the expression profile of a
 CC gene group in the sample comprising genes selected from (I). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring

CC expression levels of particular genes (AB199202 to AB199912, encoding
CC the protein sequences in AB575020 to AB575747) or by determining the
CC expression profile of a gene group comprising these genes, the
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischemic condition-improving
CC drugs or therapeutics for ischemic diseases. AB19913 and AB19914
CC represent PCR primers for a mouse ischemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 3746 BP; 1124 A; 740 C; 740 G; 1142 T; 0 other;

Query Match	93.68;	Score 23.4;	DB 24;	Length 3746;
Best Local Similarity	96.08;	Pred. No. 0.2;		
Matches 24;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

```

QY      1 CCATGTGACCATTAGGAAATGAGAG 25
         |||||
Db     501 CCATGTGACCATGAGGAATGAGAG 525

```

```

RESULT 10
AAS61690
ID      AAS61690 standard; cDNA; 3927 BP.

```

AC	AAS61690;
XX	
DT	29-JAN-2002 (first entry)
YY	

DE	Lung small cell carcinoma antigen, cDNA #231.
XX	
RW	Human: cytostatic; antitumour; lung small cell cancer antigen.
KM	tumour, lung cancer; ss.

OS	Homo sapiens.
XX	
PN	W0200177168-A2.

PD	18-OCT-2001.
XX	
PF	11-APR-2001; 2001WO-US11859.
XX	

PR 11-APR-2000; 2000US-196780P.
PR 21-JUN-2000; 2000US-213361P.
PR 01-SEP-2000; 2000US-229763P.
PR 05-SEP-2000; 2000US-230629P.
PR 14-SEP-2000; 2000US-232565P.
PR 19-DEC-2000; 2000US-257037P.
PR 08-JAN-2001; 2001US-260796P.

PA (CORI-) CORIXA CORP.

PI Lodes MJ, Wang T, Mohamath R, Indirias CY;

DR WPI; 2002-010896/01.
DR P-PSDB; AAU69409.

PT Lung tumour polynucleotide and polypeptides useful in therapy and
PT diagnosis of cancer especially lung cancer -
XX
PS Claim 1; Page 201-202; 295pp; English.

Claim 1; Page 201-202; 295pp; English.

CC The invention relates to novel isolated lung small cell cancer antigen
CC polynucleotides (I) and polypeptides (II) used in a method of detecting
CC cancer in a patient. The method is optionally performed by
CC utilizing oligonucleotides (III), where the biological sample
CC from the patient is contacted with (III), detecting the amount of
CC polynucleotide hybridized to (III) in the sample and comparing the
CC amount of polynucleotide to a predetermined cut-off value and thereby
CC determining cancer in a patient. (I), (II) or antigen-presenting cells
CC expressing (II) is useful for stimulating and/or expanding T cells
CC specific for a tumour protein. The method comprises contacting T cells
CC with one of the components under conditions to permit the stimulation
CC and/or expansion of the cells. A composition comprising (I) is useful for

CC stimulating an immune response in a patient and for inhibiting the
CC development of a cancer especially lung cancer in a patient. An
CC isolated T cell population is useful for removing tumour cells from the
CC biological sample and for inhibiting the development of cancer in a
CC patient. MS61160-MS61874 represent novel human lung small cell
CC cancer antigen coding sequences of the invention.

Query Match	93.6%	Score 23.4	DB 24	Length 3927
Best Local Similarity	96.0%	Pred. No. 0.21		
Matches 24	Conservative	0	Mismatches 1	Indels 0
Gaps				0

```

QY      1 CCATGTGACCATTAGGAATGAGAG 25
          |||||
Db      670 CCATGTGACCATGAGGAATGAGAG 694

```

```

RESULT 11
AAAX58980
ID      AAX58980 standard; cDNA; 3933 BP.

```

AC	AAx58980;
XX	
DT	23-AUG-1999 (first entry)

DE	Human transcription regulator MOP1 cDNA.
XX	
KW	MOP1, member of the PAS superfamily; bHLH-PAS; human;
KM	transcription regulator; hypoxia inducible factor 1 alpha; ss

	Homo sapiens.	Location/Qualifiers
OS		
XX		
FH	Key	85-07

FT
XX
PN

W09928464-A2.

PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98WO-US25314.
XX	
PR	28-NOV-1997; 97US-0066863.

PA (WISC) WISCONSIN ALUMNI RES FOUND.

PI Bradfield CA, Gu YZ, Hogenesch JB;
.....

DR WPI; 1999-371120/31.
DR P-PSDB; AAY06289.

PT Developmental signal transduction associated proteins
XX
PS Example 1; Page 93-94; 106pp; English.
..

CC This is the nucleotide sequence of a cDNA encoding MOP1 (see
CC AA006289), a member of the PAS superfamily, where PAS stands for
CC PER/ARNT/SIM domains. The cDNA was identified in an iterative
CC search of human ESTs designed to identify basic-helix-loop-helix-PAS
CC (bHLH-PAS) proteins that interact with either the Ah receptor (AHR)
CC or the Ah receptor nuclear translocator (ARNT). To obtain extended
CC open reading frames for each EST, an anchored-PCR strategy was used
CC to amplify additional flanking sequences from a commercial Hsp62
CC library. MOP1 is also known as hypoxia inducible factor 1 alpha.
CC The factor is induced by low oxygen. It interacts with HSP90 and
CC with ARNT. The ARNT-dimerised factor regulates expression of
CC erythropoietin, among other genes. The invention also provides
CC novel MOPs 2-9 nucleic acids (see AA058981-88) and proteins (see
CC AA06289-97). These are useful in a variety of research,
CC diagnostic and therapeutic applications. Several of the MOPs are
CC alpha-class hypoxia-inducible factors. Others are involved in
CC circadian signal transduction.

XX Sequence 3933 BP; 1243 A; 784 C; 750 G; 1156 T; 0 other;
SQ Query Match 93.6%; Score 23.4; DB 20; Length 3933;
Best Local Similarity 96.0%; Pred. No. 0.21;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CCATGTGACCATTTAGGAATGAGAG 25
|||||
Db 676 CCATGTGACCATTTAGGAATGAGAG 700

RESULT 12
AAS85058
ID AAS85058 standard; cDNA; 4162 BP.
XX AAS85058;
AC AAS85058;
XX 13-FEB-2002 (first entry)
DT
XX DNA encoding novel human diagnostic protein #20862.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR P-PSDB; ABG20871.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 20862; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pubd/published_pcc_sequences.
XX
XX Sequence 4162 BP; 1286 A; 843 C; 813 G; 1220 T; 0 other;

XX Query Match 93.6%; Score 23.4; DB 23; Length 4162;
SQ Query Match 96.0%; Pred. No. 0.21;
Best Local Similarity 96.0%; Pred. No. 0.24;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CCATGTGACCATTTAGGAATGAGAG 25
|||||
Db 804 CCATGTGACCATTTAGGAATGAGAG 828

RESULT 13
AAS14156
ID AAS14156 standard; DNA; 10355 BP.
XX AAS14156;
AC AAS14156;
XX 18-DEC-2001 (first entry)
DT
XX PSMARF CMV-HIF DNA from a vector expressing HIF-1 alpha.
DE
XX Differential expression; polymorphism; biological pathway; IRS; GFP; ds;
KM internal ribosome entry site; green fluorescent protein; HIF-1 alpha;
KM hypoxia inducible factor 1 alpha; hypoxia regulated gene; macrophage;
KM human; CMV; cytomegalovirus.
XX Synthetic.
OS
XX WO200162965-A2.
PN
XX 30-AUG-2001.
PD
XX 22-FEB-2001; 2001WO-GB00758.
PF
XX 22-FEB-2000; 2000GB-0004197.
PR 28-JUL-2000; 2000GB-0018679.
XX (OXFO-) OXFORD BIOMEDICA UK LTD.
PA
XX Kingsman AJ;
PI
XX WPI; 2001-589807/66.
DR
XX Screening a genetic element involved in a cellular process, comprises
PT comparing gene expressions in a cell, and a second cell that has
PT altered levels of genes used in the process, and detecting an element
PT with varied expression -
XX
XX Example 5; Page 97-101; 103pp; English.

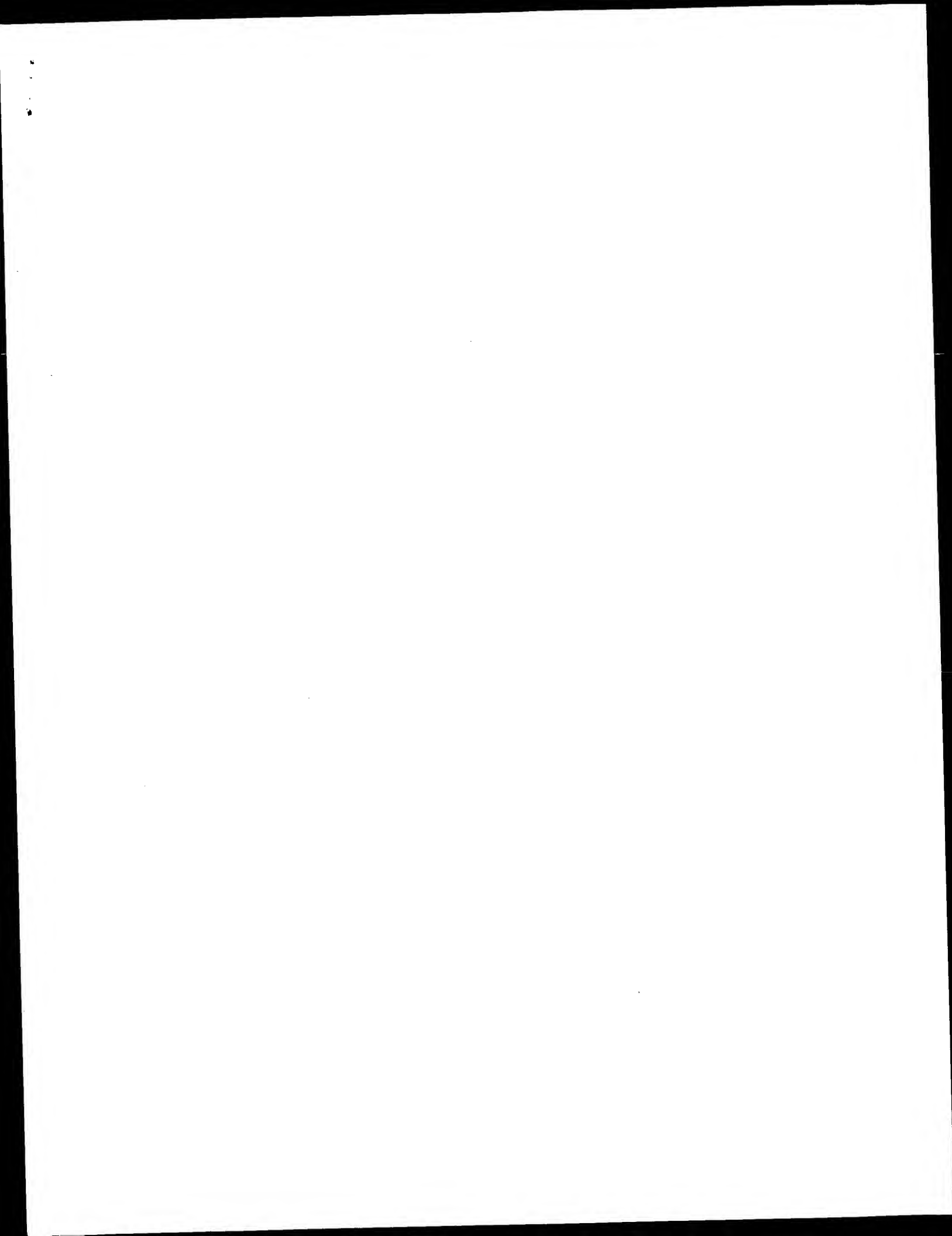
CC The invention relates to differential expression screening to identify a
CC genetic element involved in a cellular process. The method involves
CC comparing gene expressions in two cells, where one cell has altered
CC levels of a biological molecule, and identifying the genetic element
CC whose expression differs, by comparing expression under different
CC environmental conditions. The method is useful for identifying mutations
CC and polymorphisms that affect the biological response to a particular
CC cellular process. The method also allows the molecular dissection of
CC biological pathways by altering a particular pathway under study. By
CC introducing a heterologous nucleic acid into a cell population, the level
CC of a particular molecule can be influenced, allowing a pathway to be
CC dissected into its precise molecular components. The main use of the
CC method is to compare gene expression in cells from a diseased patient and
CC from a normal patient. This sequence represents PSMARF CMV-HIF from a
CC vector expressing hypoxia inducible factor 1 alpha (HIF-1 alpha) which is
CC used in methods of the invention.
XX
XX Sequence 10355 BP; 3060 A; 2212 C; 2373 G; 2710 T; 0 other;

Qy 1 CCATGTGACCATTTAGGAATGAGAG 25
Query Match 93.6%; Score 23.4; DB 22; Length 10355;
Best Local Similarity 96.0%; Pred. No. 0.24;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0251988.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 PS
 PS Disclosure: SEQ ID NO 32593; 3071bp + Sequence listing; English.
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91991. (I) have cytosolic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87994 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM2169
 CC represent sequences used in the exemplification of the present invention.
 CC
 SQ Sequence 27884 BP; 9296 A; 5084 C; 5260 G; 8244 T; 0 other;
 Query Match 93.6%; Score 23.4; DB 22; Length 27884;
 Best Local Similarity 96.0%; Pred. No. 0.27;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CCATGTGACCATTTAGGAATGAGAG 25
 DB 26482 CCATGTGACCATTTAGGAATGAGAG 26458
 RESULT 15
 AAM93977/c
 ID AAM93977 standard; cDNA: 367 BP.
 AC AAM93977;
 XX
 DT 05-OCT-2001 (first entry)
 XX
 DE Human foetal cDNA, SEQ ID NO: 506.
 XX
 KW Human; foetal protein; cytosolic; immunosuppressive; immunostimulant;
 KW neurotropic; neuroprotective; thrombolytic; osteopetrotic; antineoplastic;
 KW gene therapy; antisense therapy; cancer; immune disorder;
 KW growth disorder; osteoporosis; thrombolytic disorder;
 KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.
 XX

OS Homo sapiens.
 XX
 PN WO200155339-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001MO-US02723.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 15-SEP-2000; 2000US-0663870.
 PR 06-NOV-2000; 2000US-0707351.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Yeung G, Ford JE, Boyle BJ, Atterburn MC, Drmanac RA, Tang YT;
 PI Liu C, Asundi V, Zhou P, Werthman T;
 DR WPI: 2001-465571/50.
 DR P-PSDB; AAM06302.
 XX
 PT Novel fetal proteins useful for the treatment and diagnosis of diseases
 PT associated with dysfunction of the protein e.g. cancers, immune
 PT disorders, growth disorders, thrombolytic disorders, nervous system
 PT disorders and inflammation -
 PS
 PS Claim 1: Page 362-363; 715pp; English.
 CC
 CC The invention relates to novel foetal polypeptides encoded by
 CC polynucleotides comprising one of 477 sequences fully defined in the
 CC specification. The foetal polynucleotides and polypeptides are
 CC useful in the treatment and diagnosis of diseases such as cancers,
 CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic
 CC disorders, nervous system disorders and inflammation. The present
 CC sequence was assembled using an expressed sequence tag (EST) found
 CC to be expressed in human foetal tissue cDNA libraries as the seed.
 CC
 SQ Sequence 367 BP; 94 A; 65 C; 58 G; 150 T; 0 other;
 Query Match 72.8%; Score 18.2; DB 22; Length 367;
 Best Local Similarity 87.0%; Pred. No. 44;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 ATGTGACCATTTAGGAATGAGAG 25
 DB 219 ATGTGACCATTTAGGAATGAGAG 197

Search completed: March 17, 2003, 10:50:36
 Job time: 151.446 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 883.737 Seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-2

Sequence: 1 ccagtgaccattgagaaatgagag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthm:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: em_hic1:*
10: em_hic2:*
11: em_hic3:*
12: em_hic4:*
13: em_hic5:*
14: em_hic6:*
15: em_hic7:*
16: em_hic8:*
17: em_hic9:*
18: em_hic10:*
19: em_hic11:*
20: em_hic12:*
21: em_hic13:*
22: em_hic14:*
23: em_hic15:*
24: em_hic16:*
25: em_hic17:*
26: em_hic18:*
27: em_hic19:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	23.4	93.6	337	BQ308862 MR0-BT550
2	23.4	93.6	340	AM504525 UT-HF-BNO
3	23.4	93.6	380	AM120791 UT-M-BH2
4	23.4	93.6	386	AM124493 UT-M-BH2
5	23.4	93.6	395	BE980521 UT-M-BG2
6	23.4	93.6	501	BF406698 UT-R-BJ2

C	7	23.4	93.6	591	9	AI481068	AI481068 v192d03.x
	8	23.4	93.6	684	14	BM834484	BM834484 K-EST0109
	9	23.4	93.6	693	10	BI199902	BI199902 BI199902
	10	23.4	93.6	711	12	BE731662	BE731662 601567048
	11	23.4	93.6	734	12	BF789080	BF789080 602104940
	12	23.4	93.6	739	13	BI413382	BI413382 602986733
	13	23.4	93.6	769	12	BG772697	BG772697 602720844
	14	23.4	93.6	791	12	BG110077	BG110077 602279683
	15	23.4	93.6	878	9	AI132294	AI132294
	16	23.4	93.6	878	14	BO711945	BO711945 AGENCOURT
	17	23.4	93.6	889	14	BO213144	BO213144 AGENCOURT
	18	23.4	93.6	893	12	BE877244	BE877244 601485172
	19	23.4	93.6	941	14	BO881947	BO881947 AGENCOURT
	20	23.4	93.6	1024	14	BM807153	BM807153 AGENCOURT
	21	23.4	93.6	1657	11	AK017853	AK017853 mus muscu
	22	22.4	89.6	714	9	AI131767	AI131767
	23	21.8	87.2	310	10	BB204614	BB204614
	24	21.8	87.2	903	9	AL698195	AL698195 DKFZP6860
	25	20.4	81.6	533	17	AZ786928	AZ786928 2M0032N05
	26	19.8	79.2	372	14	W86995	W86995 z61e02.s1
	27	19.8	79.2	478	9	AA700482	AA700482 z174f07.s
	28	19.8	79.2	1435	14	BM906863	BM906863 AGENCOURT
	29	19.2	76.8	105	14	BO308421	BO308421 MR0-BT450
	30	19.2	76.8	379	14	T31654	T31654 EST36504 Hu
	31	19.2	76.8	415	17	AO108979	AO108979 CIT-HSP-2
	32	19.2	76.8	611	17	B90780	B90780 CIT-HSP-2
	33	18.8	75.2	381	17	AZ263188	AZ263188 RPT-23-1
	34	18.8	75.2	632	14	BO397883	BO397883 NISC mo02
	35	18.8	75.2	633	10	BB652302	BB652302
	36	18.8	75.2	800	12	BF617665	BF617665 HSMC001
	37	18.6	74.4	260	17	AZ778470	AZ778470 2M0013K17
	38	18.6	74.4	335	13	BJ198294	BJ198294 BJ198294
	39	18.6	74.4	366	17	AZ433190	AZ433190 1M0218L22
	40	18.6	74.4	377	13	BJ195711	BJ195711 BJ195711
	41	18.6	74.4	425	17	AZ931901	AZ931901 474.dh291
	42	18.6	74.4	468	17	CNS02728	AL184121 Telradon
	43	18.6	74.4	484	9	AA203974	AA203974 mu28g05.r
	44	18.6	74.4	519	13	BM482542	BM482542 535462 MA
	45	18.6	74.4	538	17	BH026082	BH026082 RPT-24-3

ALIGNMENTS

RESULT 1
LOCUS BQ308862 337 bp mRNA linear EST 16-MAY-2002
DEFINITION MR0-BT5505-040701-003-c02 BT5505 Homo sapiens CDNA, mRNA sequence.
ACCESSION BQ308862
VERSION BQ308862.1 GI:20851208
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 337)
Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Britons,M.R.,
Nagai,M.A., de Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bais,G.S., Simpson,D.H.,
Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:
 (http://www.ludwig.org.br/scripts/gethtml2.pl?file=MR06t2-MR0-BT5505-040701-003-c02&t3=2001-07-04&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 15
 High quality sequence stop: 331.
 Location/Qualifiers

FEATURES

source

1. 337
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT5505"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site:1; Sma1; Site:2; Sma1: A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 105 a 49 c 81 g 102 t
 ORIGIN

Query Match 93.6%; Score 23.4; DB 14; Length 337;
 Best Local Similarity 96.0%; Pred. No. 2.7;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCATGTGACCATTTAGGAATGAGAG 25
 ||||||||| |||||||||

Db 259 CCATGTGACCATTTAGGAATGAGAG 283

RESULT 2 348 bp mRNA linear EST 02-MAR-2000
 AM504525
 LOCUS
 DEFINITION
 IMAGE:3079736 5', mRNA sequence.
 AM504525
 VERSION
 AM504525.1 GI:7142192
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 NIH-MGC http://mgs.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares lab
 cDNA Library Arrayed by: M.B. Soares lab
 DNA Sequencing by: M.B. Soares lab
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbp/image/image.html
 Seq primer: M13 Forward.

FEATURES

source

Location/Qualifiers
 1. 348
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3079736"
 /clone_lib="NIH MGC_50"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LT1)"
 /note="Vector: pT73-Pac; Site:1; NotI; Site:2; Eco RI;
 constructed from size fractionated cytoplasmic mRNA
 (3.5-4.4kb). Directionally cloned. Cells provided by

Louis M. Staudt, Ph.D. Library preparation by Maria de
 Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D. "

BASE COUNT

107 a 54 c 81 g 106 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 348;
 Best Local Similarity 96.0%; Pred. No. 2.7;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCATGTGACCATTTAGGAATGAGAG 25
 ||||||||| |||||||||

Db 255 CCATGTGACCATTTAGGAATGAGAG 279

RESULT 3 380 bp mRNA linear EST 22-OCT-1999
 AM120791/c
 LOCUS
 DEFINITION
 IMAGE:3079736 5', mRNA sequence.
 AM120791
 VERSION
 AM120791.1 GI:6096124
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 house mouse.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 380)
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov

The sequence contained an oligo-dT track that was present in the strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized basal ganglia library cDNA library Preparation: M.B. Soares lab clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined.
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source

Location/Qualifiers
 1. 380
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_image="UI-M-BH2.3-any-h-05-0-UI"
 /clone_lib="NIH BMAP M.S3.3"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site:1; Not I; Site:2; Eco RI; The NIH_BMAP_M.S3.3 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M.S3.3, NIH_BMAP_M.S2, NIH_BMAP_M.S1. The subtracted library (NIH_BMAP_M.S3.3) was constructed

as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the NIH_BMAP_M.S2 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M.S3.1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996).

Research 6:791-806, 1996

TAG-LIB=NIH_BMAP_M.S3.1
TAG-TISSUE=basal-ganglia
TAG_SEQ=GTGAG

BASE COUNT 105 a 87 c 67 g 120 t 1 others

Query Match 93.6%; Score 23.4; DB 10; Length 380;
Best Local Similarity 96.0%; Pred. No. 2.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CCATGTGACCATAGGAATGAGAG 25
|||||
Db 72 CCATGTGACCATAGGAATGAGAG 48

RESULT 4

AM124493

LOCUS

386 bp mRNA linear EST 22-OCT-1999

DEFINITION UI-M-BH2.1-9po-a-05-0-UI s1 NIH_BMAP_M.S3.1 Mus musculus cDNA clone

ACCESSION

AM124493

VERSION

AM124493.1

KEYWORDS

GI:6099988

SOURCE

EST

ORGANISM

house mouse.

REFERENCE

Bonaldo, M.F., Lennon, G. and Soares, M.B.

AUTHORS

1 (bases 1 to 386)

TITLE

Normalization and subtraction: two approaches to facilitate gene

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

COMMENT

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mestr@nhi.nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

normalized amygdala library cDNA library preparation: M.B. Soares

Lab Clone distribution: NIH BMAP cDNA clones will be made available

by the means that is soon to be determined. When NIH determines the

means for distribution of the BMAP cDNA clones, this record will be

updated accordingly when that means is determined.

Seq primer: M13 Forward

POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..386

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UI-M-BH2.1-9po-a-05-0-UI"

/clone_lib="NIH_BMAP_M.S3.1"

/dev_stage="27-32 days"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

BASE COUNT 106 a 72 c 91 g 117 t

ORIGIN

Query Match

Best Local Similarity 96.0%; Score 23.4; DB 10; Length 386;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CCATGTGACCATAGGAATGAGAG 25
|||||
Db 315 CCATGTGACCATAGGAATGAGAG 339

RESULT 5

BE980521

LOCUS

395 bp mRNA linear EST 05-OCT-2000

DEFINITION UI-M-BG2-bck-b-03-0-UI s1 NIH_BMAP_M.S3.1 Mus musculus cDNA clone

ACCESSION

BE980521

VERSION

BE980521.1

KEYWORDS

GI:10648591

SOURCE

EST

ORGANISM

house mouse.

REFERENCE

Bonaldo, M.F., Lennon, G. and Soares, M.B.

AUTHORS

1 (bases 1 to 395)

TITLE

Normalization and subtraction: two approaches to facilitate gene

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

COMMENT

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mestr@nhi.nih.gov

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to verify it as a clone from the

amygdala tissue cDNA library preparation: M.B. Soares Lab Clone

distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and

PolyLinker: Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_M.S3.1 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged, normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following process: NIH_BMAP_M.S3.1, NIH_BMAP_M.S2, NIH_BMAP_M.S1. The subtracted library (NIH_BMAP_M.S3.1) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the NIH_BMAP_M.S2 library single-stranded circles (subtracted library). The remaining by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M.S3.1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996).

TAG-LIB=NIH_BMAP_M.S3.1
TAG-TISSUE=amygdala
TAG_SEQ=GTGAG

limited collaborative arrangements
Seq primer: M13 Forward
POLYA-Yes.

FEATURES

source

Location/Qualifiers
1. 395

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BG2-bck-b-03-0-UI"
/clone_lib="NIH_BMAP_MSC_S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_MSC_S1 library is a subtracted library derived
from NIH_BMAP_MSC_N. NIH_BMAP_MSC_N was made from mouse spinal
cord tissue. For a detailed description of the library
from which this clone was derived, please visit our web
site at brainest.eng.uiowa.edu.
TAG_LIB="NIH_BMAP_MSC_S1"
TAG_TISSUE="amygdala"
TAG_SEQ="GTGAG"

BASE COUNT 108 a 73 c 93 g 121 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 12; Length 395;
Best Local Similarity 96.0%; Pred. No. 2.9;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CCATGTGACCATTTAGGAATGAG 25

DB 315 CCATGTGACCATTTAGGAATGAG 339

RESULT 6 501 bp mRNA linear EST 28-NOV-2000

LOCUS

BF406698 UI-R-BJ2-bpo-b-02-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone

DEFINITION UI-R-BJ2-bpo-b-02-0-UI 3', mRNA sequence.

ACCESSION BF406698

VERSION BF406698.1

KEYWORDS GI:11394673

EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 501)

Bonaldi, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Oligo-dT track not found. Not a site shown in beginning of sequence

is likely internal to the message. cDNA Library Preparation: M.B.

Soares Lab Clone distribution: clones will be available through

Research Genetics (www.resgen.com) The following repetitive

elements were found in this cDNA sequence: 25-96, >B1-F#SINE/Alu

Seq primer: M13 Forward

POLYA-No.

Location/Qualifiers

1. 501

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-BJ2-bpo-b-02-0-UI"

/clone_lib="UI-R-BJ2"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2

library is a subtracted library derived from the following

tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,

atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15

dpc, AV canal at 15 dpc. For a detailed description of

the library from which this clone was derived, please

visit our web site at brainest.eng.uiowa.edu. The

subtraction has been previously described in (Bonaldi,

Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

TAG_SEQ="None found"

BASE COUNT 150 a 143 c 116 g 182 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 9; Length 591;
Best Local Similarity 96.0%; Pred. No. 3.4;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CCATGTGACCATAGGAATGAGAG 25
|||||
Db 390 CCATGTGACCATAGGAATGAGAG 366

RESULT 8
BM834484

LOCUS BM834484 684 bp mRNA linear EST 06-MAR-2002
DEFINITION K-EST0109485 S11SNUI Homo sapiens CDNA clone S11SNUI-63-B11 5',
BM834484
VERSION BM834484.1 GI:19190893
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
1 (bases 1 to 684)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Institute of Bioscience & Biotechnology
52 Deoun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsungemail.kr@ib.re.kr
Plate: 63 row: B column: 11
High quality sequence stop: 684.
Location/Qualifiers

FEATURES
source

1. 684
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S11SNUI-63-B11"
/clone_lib="S11SNUI"
/sex="M"
/tissue_type="Stomach"
/cell_line="Lymphoblast-like"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pME18-FL3; Site: 1: XhoI;
Site: 2: XhoI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deprotected
with tobacco acid pyrophosphatase (TAP). The deprotected
intact mRNA was ligated with DNA-RNA linker including SfiI
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using SfiI
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with SfiI and
cloned into DraIII-digested pME18-FL3 vector. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 232 a 135 c 136 g 181 t

Query Match 93.6%; Score 23.4; DB 14; Length 684;
Best Local Similarity 96.0%; Pred. No. 3.6;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CCATGTGACCATAGGAATGAGAG 25
|||||
Db 75 CCATGTGACCATAGGAATGAGAG 99

RESULT 9
BB199902

LOCUS BB199902 693 bp mRNA linear EST 19-OCT-2001
DEFINITION BB199902 RIKEN full-length enriched, 0 day neonate thymus Mus
musculus CDNA clone A430018A13 3', mRNA sequence.
BB199902
VERSION BB199902.2 GI:16271363
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hitamoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Kono,H., Koda,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jun 30, 2000 this sequence version replaced gi:8864855.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujisaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Wahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,
S., Kawai,T., Okazaki,Y., Muramatsu,M., Inoue,Y., Kite,A. and
Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)
COMMENT Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamane,I., Aizawa,
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers

FEATURES
source

1. 693
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A430018A13"
/clone_lib="RIKEN full-length enriched, 0 day neonate
thymus"

```

/tissue.type="thymus"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: Sall; Site_2: BamHI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGAGAGATCCAGACGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGGAGAGATTTCTCGAGTTATTTAAATATATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified Bluescript KS(+) after bulk excision from Lambda

```

	Query Match	93.6%	Score 23.4;	DB 10;	Length 893;
	Best Local Similarity	96.0%;	Pred. No. 3.6;		
Matches	24; Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;
OY	1 CCATGTACCATGAGCAATGAGAG	25			
Dδ	414 CCATGTACCATGAGCAATGAGAG	438			

RESULT 10	BE731662	711 bp	mRNA	linear	EST 15-SEP-2000
LOCUS	BE731662				
DEFINITION	BE731662	711 bp	mRNA	linear	EST 15-SEP-2000
	601567048F				
	NH_MGC_21 Homo sapiens cDNA clone IMAGE:5842146 5', mRNA sequence.				

ACCESSION	BE731662	GI:10145654
VERSION	BE731662.1	
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 711)
TITLE	NH-MGC http://mgc.nci.nih.gov/
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lln.gov
Plate: L10M534 row: o column: 11
High quality sequence stop: 697.
Location/Qualifiers

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3842146"
/clone_1lb="NTH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pONB7; Site_1: XhoI;
Site_2: EcoRI. cDNA made by oligo-dT priming,
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size selected >500bp
for average insert size 1.8kb. Library constructed by
Gerald M. Rubin (University of California, Berkeley) using
ZAP-cDNA synthesis kit

```

BASE COUNT	174 a	166 c	197 g	174 t
ORIGIN	(Stratagene) and Superscript II RT (Life Technologies).			

Query Match	93.68;	Score 23.4;	DB 12,	Length 711;
Best Local Similarity	96.0%;	Pred. No. 3.6;		
Matches 24;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

QY 1 CCATGTGACCATTAGGAATGAGAG 25
 |||||
 Db 646 CCATGTGACCATTAGGAATGAGAG 670

RESULT 11					
BE789080					
LOCUS	BE789080	734 bp	mRNA	linear	EST 12-JAN-2001
DEFINITION	602104940F1 NCL- <i>CGAP_Kid14</i> Mus musculus cDNA clone IMAGE:4223096				
	5', mRNA sequence.				

ACCESSION	BF789080
VERSION	BF789080.1
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

ORGANIZATION: National Institutes of Health
 EDUCATION: Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
 1 (base) 1 to 734
 NIH-MGC: <http://mgc.nci.nih.gov/>
 NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
 UNPUBLISHED (1999)
 CONTACT: Robert Strausberg, Ph.D.
 CONTACT: rob@b-remail.nih.gov

Llama: 3000
 Tissue Procurement: Jeffrey E. Green M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MCC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM9810 row: p column: 09
 High quality sequence stop: 708.

FEATURES
source

```

SOURCE
/organism="Mus_musculus"
/strain="FVB/N"
/db_xref="tacon:10090"
/clone_1 IMAGE:4223096"
/clone_1b="NCI_CGAP_K1d14"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: kidney; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT
217 a 147 c 170 g 200 t

```

Query Match

Query Match	93.68;	Score 23.4;	DB 12;	Length 734;
Similarity	96.08;	Pred. No. 3.7;		
Best Local	0;	Mismatches	1;	Indels 0;
Matches	24;	Conservative		Gaps 0;

QY 1 CCATGTGACCATTAGGAATGAGAG 25
 |||||
 Db 306 CCATGTGACCATTAGGAATGAGAG 330

RESULT 12				
B1413382				
LOCUS	B1413382	739 bp	mRNA	linear
DEFINITION	6029667373F1 NCL_CGAP_Lu33	MMS	MUSCULUS	CDNA
				clone
				IMAGE:1442800 5',
				mRNA sequence.

ACCESSION	BI413382
VERSION	BI413382.1
KEYWORDS	EST.
SOURCE	house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 739)
NIH-MGC <http://imgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM11352 row: a column: 09
High quality sequence start: 125
High quality sequence stop: 726.
Location/Qualifiers

FEATURES

source

1. 739
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5142800"
/clone_lib="NCI-CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pRT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTTACCAATCTAGAGTGGAGCGCCCTCTCTTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

180 a 176 c 210 g 171 t 2 others

Query Match

Best Local Similarity 93.6%; Score 23.4; DB 13; Length 739;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 CCATGTGACCATTTAGGAATGAGAG 25
|||||
681 CCATGTGACCATTTAGGAATGAGAG 705

RESULT 13

LOCUS BG772697 769 bp mRNA linear EST 15-MAY-2001
DEFINITION 602720844F1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4837691 5',
mRNA sequence.

ACCESSION BG772697.1 GI:14083350
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 769)
NIH-MGC <http://imgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Niklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shih-
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

FEATURES

source

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM10770 row: h column: 12
High quality sequence stop: 723.
Location/Qualifiers
1. 769
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4837691"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
size-selected for average insert size 2.2 kb and
normalized to ROP 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI), National
Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT

243 a 143 c 170 g 213 t

Query Match

Best Local Similarity 93.6%; Score 23.4; DB 12; Length 769;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 CCATGTGACCATTTAGGAATGAGAG 25
|||||
434 CCATGTGACCATTTAGGAATGAGAG 458

RESULT 14

LOCUS BG110077 791 bp mRNA linear EST 30-JAN-2001
DEFINITION 602279683F1 NIH_MGC_86 Homo sapiens CDNA clone IMAGE:4367229 5',
mRNA sequence.

ACCESSION BG110077
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 791)
NIH-MGC <http://imgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM10019 row: e column: 22
High quality sequence stop: 715.
Location/Qualifiers
1. 791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4367229"
/clone_lib="NIH_MGC_86"
/tissue_type="osteosarcoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bone; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.533 kb. Library enriched for
full-length clones and constructed by Life Technologies."

FEATURES

source

Note: this is a NIH_MGC Library."

BASE COUNT 250 a 145 c 173 g 223 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 12; Length 791;
Best Local Similarity 96.0%; Pred. No. 3.8;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCATGTGACCATAGGAATGAGAG 25
|||||

Db 236 CCATGTGACCATAGGAATGAGAG 260

RESULT 15

AU132294 878 bp mRNA linear EST 01-AUG-2002
LOCUS AU132294 NT2RP3 Homo sapiens cDNA clone NT2RP3004165 5', mRNA
DEFINITION sequence.

ACCESSION AU132294
VERSION AU132294.1 GI:10992648

KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 878)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.

REFERENCE

AUTHORS
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

source

1..878
Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_id="NT2RP3004165"

/clone_lib="NT2RP3"

/cell_type="teratocarcinoma"

/cell_line="NT2"

/note="Vector: pMT18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

BASE COUNT 274 a 163 c 188 g 248 t 5 others

ORIGIN

Query Match 93.6%; Score 23.4; DB 9; Length 878;
Best Local Similarity 96.0%; Pred. No. 4;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCATGTGACCATAGGAATGAGAG 25
|||||

Db 375 CCATGTGACCATAGGAATGAGAG 399

Search completed: March 17, 2003, 13:09:12
Job time : 888.737 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 220.903 Seconds
(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-3

Perfect score: 24

Sequence: 1 gcttcttgcacagagcgcgca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GeneBank:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.4	89.2	175	AF249160	AF249160 Philautus
2	21.4	89.2	1056	PMY18677	Y18677 Sardinia p11
3	21.4	89.2	2577	CIRH11	AF149320 Columba 1
4	21.4	89.2	3016	I18746	I18746 Sequence 1
5	21.4	89.2	5385	AB065668	AB065668 Homo sapi
6	21.4	89.2	6953	I18747	I18747 Sequence 2
7	21.4	89.2	6953	HS049742	U49742 Human rhodo
8	21.4	89.2	163297	AC023162	AC023162 Homo sapi
9	21.4	89.2	164396	HS049742	AC000380 Homo sapi
10	21.4	89.2	168551	AC080007	U97272 Abyssococci
11	20.4	85.0	867	AF249130	AF249130 Hyperolius
12	19.8	82.5	175	AF249133	AF249133 Mantella
13	19.8	82.5	175	AF249134	AF249134 Mantella
14	19.8	82.5	175	AF249136	AF249136 Boophis x
15	19.8	82.5	175	AF249137	AF249137 Boophis t
16	19.8	82.5	175	AF249138	AF249138 Laliostom
17	19.8	82.5	175	AF249139	AF249139 Fejervary
18	19.8	82.5	175	AF249140	AF249140 Fejervary
19	19.8	82.5	175	AF249141	AF249141 Hoplobatr
20	19.8	82.5	175	AF249142	AF249142 Sphaeroth
21	19.8	82.5	175	AF249143	AF249143 Euphyictl
22	19.8	82.5	175	AF249144	AF249144 Nannophry
23	19.8	82.5	175	AF249145	AF249145 Nyctibatr
24	19.8	82.5	175	AF249147	AF249147 Limnodyn
25	19.8	82.5	175	AF249148	AF249148 Limnodyn
26	19.8	82.5	175	AF249149	AF249149 Rana curt
27	19.8	82.5	175	AF249150	AF249150 Rana temp
28	19.8	82.5	175	AF249151	AF249151 Rana temp
29	19.8	82.5	175	AF249152	AF249152 Micrixal
30	19.8	82.5	175	AF249154	AF249154 Indirana
31	19.8	82.5	175	AF249156	AF249156 Polypedat
32	19.8	82.5	175	AF249157	AF249157 Rhacophor
33	19.8	82.5	175	AF249158	AF249158 Philautus
34	19.8	82.5	726	AF369050	AF369050 Polyodon
35	19.8	82.5	858	AF137208	AF137208 Amla calv
36	19.8	82.5	864	AF137206	AF137206 Aclipsner
37	19.8	82.5	867	U97265	U97265 Coltoconeph
38	19.8	82.5	867	U97266	U97266 Coltoconeph
39	19.8	82.5	867	U97274	U97274 Comophorus
40	19.8	82.5	1047	AF055319	AF055319 Trichechu
41	19.8	82.5	1053	AF021242	AF021242 Melopsitt
42	19.8	82.5	1150	AF008847	AF008847 Sus scrof
43	19.8	82.5	1300	AF309568	AF309568 Spilax eh
44	19.8	82.5	1389	PMY18679	Y18679 Pomatoschis
45	19.8	82.5	1389	PMY18679	Y18679 Pomatoschis

ALIGNMENTS

RESULT 1
AF249160 175 bp DNA linear VRT 17-JAN-2001
LOCUS Philautus charius rhodopsin gene, exon 4 and partial cds.
DEFINITION
ACCESSION AF249160
VERSION AF249160.1 GI:12247249
KEYWORDS
SOURCE
ORGANISM Philautus charius.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rhacophoridae;
Philautus.
REFERENCE
AUTHORS 1 (bases 1 to 175)
TITLE Bossuyt, F. and Minkovitch, M.C.
Convergent Adaptive Radiations in Madagascan and Asian Rain Forest Frogs

Reveal Co-variation between Larval and Adult Traits
 Unpublished
 2 (bases 1 to 175)
 AUTHORS Bossuyt, F. and Milinkovitch, M.C.
 TITLE Direct Submission
 JOURNAL Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of Molecular Biology and Medicine, rue Jenner and Brachet 12, Gosselies B-6041, Belgium

FEATURES

source

1..175
 /organism="Philaenus charius"
 /db_xref="taxon:129025"

mrna

<1..>175
 /product="rhodopsin"

CDS

<1..>175
 /product="rhodopsin"

/protein_id="AAG49803.1"

/db_xref="GI:12247250"

/translation="AEKEVTRWVIMVVEFLICWPAVAVAFYFTHGSSEFGPIFMT
 VPAPFAKSSATYNP"

exon

<1..>175
 /number=4

BASE COUNT

32 a 55 c 36 g 52 t

ORIGIN

Query Match 89.2%; Score 21.4; DB 5; Length 175;
 Best Local Similarity 95.7%; Pred. No. 5.5;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCCAGAGCGCCGC 23

Db 140 GCTTCTTTGCCAGAGCTCCGC 162

RESULT 2

SPY18677

LOCUS

Sardina pilchardus mRNA for opsin.

DEFINITION

Y18677

ACCESSION

Y18677.1

GI:4210840

KEYWORDS

opsin.

Sardina pilchardus.

Sardina pilchardus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Clupeidae;

Sardina.

1 (bases 1 to 1056)

Archer, S.N. and Hirano, J.

Comparative analysis of opsins in Mediterranean coastal fish

Unpublished

2 (bases 1 to 1056)

Archer, S.N.

Direct Submission

Submitted (25-JAN-1999) S.N. Archer, International Marine Centre,

Localita sa Mardini, 09072 Torregrande, Oristano, ITALY

Location/Qualifiers

1..1056

/organism="Sardina pilchardus"

/db_xref="taxon:27697"

/tissue="retina"

/dev_stage="adult"

1..1056

/function="phototransduction"

/codon_start=1

/product="opsin"

/protein_id="CAA77259.1"

/db_xref="GI:4210841"

/translation="MNGEGPEFYIPMSNATGLVRSPPDYPOYTLVPPMGYACLAAYV FLILITGPVNPFLITLYTIEHKILRSPILNLAVALDFMVGFTTMTSLNGY FVFGMGCNIGFPTLGGETALNSIVLSEKRIYVCKPISNRRGENHAYMGVAAS WPMACAVPPLVGNRSRIIPGMOCSCGIDYTTREGNNNSFYIMVAFVHFTCLPLI ITFCYGRIVCTVEKAAAOQDESETTQRAERIVTIVIMFVAFLACWVPAVSAMVYIF

THOGESEGPVEMTIPAFPAKSSAVNPVYIICLNKQFRHQMITTLCCGKNPPEEERGS
 TRASKTEASSVCVSPA"

misc_feature

1..17

/note="degenerate primer"

1046..1056

/note="degenerate primer"

368 c 269 g 233 t

BASE COUNT 186 a

368 c 269 g 233 t

ORIGIN

Query Match 89.2%; Score 21.4; DB 5; Length 1056;
 Best Local Similarity 95.7%; Pred. No. 6.6;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCCAGAGCGCCGC 23

Db 874 GCTTCTTTGCCAGAGCTCCGC 896

RESULT 3

CLRH1

LOCUS

Columbia livia RHL opsin (rh1) gene, exons 1 through 4.

AF149230

AF149230.1 GI:4887218

KEYWORDS

1 of 2

Columbia livia.

Columbia livia

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.

1 (bases 1 to 2577)

Kawamura, S., Blow, N.S. and Yokoyama, S.

Genetic analyses of visual pigments of the pigeon (Columbia livia)

Genetics 153 (4), 1839-1850 (1999)

MEDLINE 20050679

PUBMED 10581289

REFERENCE 2 (bases 1 to 2577)

Yokoyama, S.

Direct Submission

Submitted (10-MAY-1999) Department of Biology, Syracuse University,

130 College Place, Syracuse, NY 13244, USA

Location/Qualifiers

1..2577

/organism="Columbia livia"

/db_xref="taxon:8932"

<76..436

/gene="rh1"

/number=1

1293..1461

/gene="rh1"

/number=2

1561..1726

/gene="rh1"

/number=3

1968..2207

/gene="rh1"

/number=4

442 a 778 c 815 g 542 t

BASE COUNT

ORIGIN

Query Match 89.2%; Score 21.4; DB 5; Length 2577;
 Best Local Similarity 95.7%; Pred. No. 7.1;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCCAGAGCGCCGC 23

Db 2145 GCTTCTTTGCCAGAGCTCCGC 2167

RESULT 4

LOCUS

118746

Sequence 1 from patent US 5498521.

118746

3016 bp

DNA

linear

PAT 07-OCT-1996

```
ACCESSION 118746
VERSION 118746.1 GI:1599101
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 3016)
TITLE Dryja,T.P. and Berson,E.L.
JOURNAL Diagnosis of hereditary retinal degenerative diseases
FEATURES
BASE COUNT 689 a 863 c 753 g 711 t
ORIGIN
Query Match
Best Local Similarity 89.2%; Score 21.4; DB 6; Length 3016;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCTTCTTTGCCAAGAGCGCCG 23
Db 118746 GCTTCTTTGCCAAGAGCGCCG 1190

RESULT 5
LOCUS AB065668
DEFINITION Homo sapiens gene for seven transmembrane helix receptor, complete
ACCESSION AB065668
VERSION AB065668.1 GI:21928610
KEYWORDS
SOURCE
ORGANISM Homo sapiens (Isolate:CBRC7M_231) DNA.
REFERENCE
AUTHORS Suwa,M., Sato,T., Okouchi,I., Arita,M., Putani,K., Matsumoto,S.,
TITLE Tsubumi,S., Aburatani,H., Asai,K. and Akiyama,Y.
JOURNAL Genome-wide discovery and analysis of human seven transmembrane
REFERENCE Unpublished
AUTHORS Suwa,M.
TITLE 2 (bases 1 to 5385)
JOURNAL
COMMENT Direct Submission
Submitted (11-JUL-2001) Makiko Suwa, Computational Biology Research
Center (CBRC), National Institute of Advanced Industrial Science
and Technology (AIST); 2-41-6 Aomi Koto-ku, Tokyo 135-0064, Japan
(E-mail: suwa@aist.go.jp, URL: http://www.cbrc.jp/,
Tel: 81-3-3599-8080, Fax: 81-3-3599-8081)
This sequence is a seven transmembrane helix receptor candidate
system that contains programs of gene
finding(GeneDecoder), sequence search, motif-domain assignment and
transmembrane helix prediction.
And the sequence is submitted by the collaborative project between
[Computational Biology Research Center (CBRC), National Institute
of Advanced Industrial Science and Technology (AIST)] and [Genome
Science Division, Research Center for Advanced Science and
Technology (RCAST), University of Tokyo].
FEATURES
source
1. 5385
/organism="Homo sapiens"
/isolate="CBRC7M_231"
/db_xref="taxon:9606"
/chromosome="3"
join(201..561,2343..2511,3718..3883,4000..4239,5075..5185)
/codon_start=1
/evidence=not_experimental
/product="seven transmembrane helix receptor"
/protein_id="BAC05894.1"
/db_xref="GI:21928611"

CDs
1. 5385
/organism="Homo sapiens"
/isolate="CBRC7M_231"
/db_xref="taxon:9606"
/chromosome="3"
join(201..561,2343..2511,3718..3883,4000..4239,5075..5185)
/codon_start=1
/evidence=not_experimental
/product="seven transmembrane helix receptor"
/protein_id="BAC05894.1"
/db_xref="GI:21928611"

RESULT 6
LOCUS 118747
DEFINITION Sequence 2 from patent US 5498521.
ACCESSION 118747
VERSION 118747.1 GI:1599102
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 6953)
TITLE Dryja,T.P. and Berson,E.L.
JOURNAL Diagnosis of hereditary retinal degenerative diseases
FEATURES
BASE COUNT 1523 a 2022 c 1797 g 1611 t
ORIGIN
Query Match
Best Local Similarity 89.2%; Score 21.4; DB 6; Length 6953;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCTTCTTTGCCAAGAGCGCCG 23
Db 4272 GCTTCTTTGCCAAGAGCGCCG 4294

RESULT 7
LOCUS HS049742
DEFINITION Human rhodopsin gene, complete cds.
ACCESSION U49742 K02281
VERSION U49742.1 GI:1236136
KEYWORDS opsin; rhodopsin.
SOURCE
ORGANISM Homo sapiens
REFERENCE
AUTHORS Nathans,J. and Hogness,D.S.
TITLE Isolation and nucleotide sequence of the gene encoding human
rhodopsin
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (15), 4851-4855 (1984)
MEDLINE 84272729
PUBMED 6589631
REFERENCE
AUTHORS Nathans,J.
TITLE Direct Submission
JOURNAL Submitted (22-FEB-1996) Jeremy Nathans, Molecular Biology and
Genetics, Johns Hopkins Medical School, 725 N. Wolfe Street,
Baltimore, MD 21205, USA
On Sep 3, 1996 this sequence version replaced gi:189393.

/translacion="MNGEGPNFVPSNATGVRSPEYPOYLAEPMQFSLAAVY
FLIVIGPINFILYTVQHKRITPLNTILNLAVADLFVIGFTSLYSLNGY
FVKGTCNGEGFRITLGGELALMSLVLAIERVYVCKPMSNREGENALINGVAF
VWALACAPPLAGMSRYIPGSLQSGIDYVTLKPEVNNSEFVYVAFVHTIPMI
IFECYGLVFPVKEAQAQOOSATTOAKAEKVRMVIINVIATLGMVPAASVAFIF
THGSNFGPIEMTIPAFPAKSAIYNVYIIMNKOPRNCMLTTICGKRNPLGDDAS
ATVSKTEFSQVAPPA"
```


RESULT 11					
957272	957272				
LOCUS					
DEFINITION	abyssocottus korotneffi	867 bp	DNA	linear	VRT 12-AUG-1999
ACCESSION	U97272				
VERSION	U97272.1	GI:2226036			
KEYWORDS					
SOURCE	abyssocottus korotneffi.				
ORGANISM	abyssocottus korotneffi				

REFERENCE	1	(bases 1 to 867)
AUTHORS	Hunt,D.M., Fitzgibbon,J., Slobodyanyuk,S.J., Bowmaker,J.K. and Dulai,K.S.	
TITLE	Molecular evolution of the cottoid fish endemic to Lake Baikal deduced from nuclear DNA evidence	
JOURNAL	Mol. Phylogenet. Evol.	8 (3), 415-422 (1997)
MEDLINE	98086781	
PUBMED	9417898	
REFERENCE	2	(bases 1 to 867)
AUTHORS	Hunt,D.M., Fitzgibbon,J., Slobodyanyuk,S.J., Bowmaker,J.K. and Dulai,K.S.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-APR-1997)	Molecular Genetics, Ophthalmology, Bathing
FEATURES	Street, London EC1V 9EL, UK	
source	Location/Qualifiers	
	1. 867	
	/organism="Abyssocottus korotneffi"	
CDS	/db_xref="taxon:61637"	
	<1..2867	

```

/codon_start=1
/product="rod ops in"
/protein_id="AAB61726.1"
/db_xref="GI:2226031"
/translation="YLVNPAAYALGAYMFLILIGFPINFLVLTVEHKKRPLPLN
YILNLINVAALFNWLGCTFTTMYTSMHGFEVIGRCGNCNEAFATLGEIILWMSLYVI

```

Query Match	85.0%;	Score 20.4;	DB 5;	Length 867;
Best Local Similarity	95.5%;	Pred. No. 21;		
Matches 21;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
OY	2 CTTTCTTTCACAAGACCCGC	23		
Db	788 CCTTTCTTTCACAAGACCCGC	809		
BASE COUNT	152 a	293 c	223 g	199 t
ORIGIN	AATRNIVCKPSPINRPFEDHAIINGLAFETVMAIACAPPLVNGSRYPIDGMOCCGV DYTTAAEGNNSEFVIYMFIVHPLIPLSVIFCYGLLCAVEAPAAQOSETTORAE KVASRRIYAVIGFGLVCMPLPASYAMMTFCONGSDGPIFMTLPSEFFAKSAIYNPMI YICMKRHHOMIT ⁺			

RESULT 12				
AF249130				
LOCUS				
DEFINITION	AF249130	175 bp	DNA	linear
ACCESSION	Hyperolius sp.			
VERSION	AF249130			
KEYWORDS	AF249130.1	GI:12247189		
SOURCE				
ORGANISM	Hyperolius sp.			
	Hyperolius sp.			
	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Amphibia; Batrachia; Anura; Neobatrachia; Ranolae; Hyperollidae;			
	Hyperolius.			
REFERENCE	1 (bases 1 to 175)			
AUTHORS	Bossuyt,F. and Milinkovitch,M.C.			
TITLE	Convergent Adaptive Radiations in Madagascan and Asian Raind Frogs			
	Reveal Co-variation between Larval and Adult Traits			
	Unpublished			
JOURNAL	2 (bases 1 to 175)			
REFERENCE	Bossuyt,F. and Milinkovitch,M.C.			
AUTHORS	Direct Submission			
TITLE	Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of			
JOURNAL	Molecular Biology and Medicine, rue Jeener and Brichet 12,			
	Gosselies B-6041, Belgium			
FEATURES	Location/Qualifiers			
source	1..175			

[illegible]

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
mRNA
CDS

Mantella madagascariensis.
Mantella madagascariensis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rhacophoridae;
Mantella.
1 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Reveal Co-variation between larval and Adult Traits
Unpublished
2 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Direct Submission
Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
Molecular Biology and Medicine, rue Jeener and Brachet 12,
Gosselies B-6041, Belgium
Location/Qualifiers
1..175
/organism="Mantella madagascariensis"
/db_xref="taxon:111290"
/product="rhodopsin"
/codon_start=2
/product="rhodopsin"
/protein_id="AA049776.1"
/db_xref="GI:12247196"
/translation="AEKEVTRMVMVFFLICWVPAVVAFYIFTHGSEFGPIFMT
VPAFAKSSAIVNP"
/number=4
1..>175
/number=4

BASE COUNT 33 a 53 c 36 g 52 t 1 others
ORIGIN
Query Match 82.5%; Score 19.8; DB 5; Length 175;
Best Local Similarity 91.3%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCCAAGAGCGCCG 23
140 GCTTCTTTGCCAAGAGCTCTGC 162

RESULT 14
AF249134 175 bp DNA linear VRT 17-JAN-2001
LOCUS
DEFINITION
ACCESSION
AF249134
AF249134.1 GI:12247197
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
mRNA
CDS

Mantidactylus cf. ulcerosus.
Mantidactylus cf. ulcerosus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rhacophoridae;
Mantidactylus.
1 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
Reveal Co-variation between larval and Adult Traits
Unpublished
2 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Direct Submission
Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
Molecular Biology and Medicine, rue Jeener and Brachet 12,
Gosselies B-6041, Belgium
Location/Qualifiers
1..175
/organism="Mantidactylus cf. ulcerosus"
/db_xref="taxon:129014"
/product="rhodopsin"
1..>175

/codon_start=2
/product="rhodopsin"
/protein_id="AA049777.1"
/db_xref="GI:12247198"
/translation="AEKEVTRMVMVFFLICWVPAVVAFYIFTHGSEFGPIFMT
VPAFAKSSAIVNP"
1..>175
/number=4

BASE COUNT 36 a 50 c 36 g 53 t
ORIGIN
Query Match 82.5%; Score 19.8; DB 5; Length 175;
Best Local Similarity 91.3%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCCAAGAGCGCCG 23
140 GCTTCTTTGCCAAGAGCTCTGC 162

RESULT 15
AF249136 175 bp DNA linear VRT 17-JAN-2001
LOCUS
DEFINITION
ACCESSION
AF249136
AF249136.1 GI:12247201
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
mRNA
CDS

Boophis xerophilus.
Boophis xerophilus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rhacophoridae;
Boophis.
1 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Convergent Adaptive Radiations in Madagascan and Asian Ranid Frogs
Reveal Co-variation between larval and Adult Traits
Unpublished
2 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Direct Submission
Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
Molecular Biology and Medicine, rue Jeener and Brachet 12,
Gosselies B-6041, Belgium
Location/Qualifiers
1..175
/organism="Boophis xerophilus"
/db_xref="taxon:128996"
/product="rhodopsin"
/codon_start=2
/product="rhodopsin"
/protein_id="AA049779.1"
/db_xref="GI:12247202"
/translation="AEKEVTRMVMVFFLICWVPAVVAFYIFTHGSEFGPIFMT
VPAFAKSSAIVNP"
1..>175
/number=4

BASE COUNT 34 a 55 c 34 g 52 t
ORIGIN
Query Match 82.5%; Score 19.8; DB 5; Length 175;
Best Local Similarity 91.3%; Pred. No. 37;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCCAAGAGCGCCG 23
140 GCTTCTTTGCCAAGAGCTCTGC 162

Search completed: March 17, 2003, 11:25:45
Job time : 313.903 secs

FT	misc_binding		complement (355..369)
FT		/tag- f	
FT		/note= "binds probes AAT17119 (mutant) and AAT17120	
FT		(normal)"	
FT	mutation	362	
FT		/tag- g	
FT		/note= "Substitution with A in mutant sequence"	
FT	primer_bind	362..381	
FT		/tag- h	
FT		/note= "binds primers 485 (AAT17122) (normal) and 502	
FT		(mutant)"	
XX			
XX	US5498521-A.		
XX			
PD	12-MAR-1996.		
XX			
FP	24-JAN-1990;	90US-0469215.	
XX			
PR	11-MAR-1993;	93US-0033081.	
PR	24-JAN-1990;	90US-0469215.	
PR	11-DEC-1991;	91US-0805123.	
XX			
PA	(HARD) HARVARD COLLEGE.		
XX			
P1	Berson EL Dryja TP;		
DR	WPI; 1996-159684/16.		
DR	P-PSDB; AAR93116.		
XX			
PT	Diagnosis of hereditary retinal degenerative diseases e.g. retinitis		
PT	pigmentosa, caused by a human photoreceptor protein mutation, by		
PT	detection of the mutation by PCR amplification or hybridisation		
PT	methods		
XX			
PS	Example 1; Column 19-24; 71pp; English.		
XX			
CC	This sequence encodes human rhodopsin, and is shown without		
CC	introns. The full sequence, with introns, is shown in AAT17116.		
CC	Substitution of histidine for the normal nonpolar amino acid		
CC	proline at position 23, by substitution of C with A in codon -23,		
CC	results in a dysfunctional or absent molecule, affecting rod		
CC	function, and is linked with autosomal dominant retinitis		
CC	pigmentosa. Probes AAT17117 and AAT17119 bind to the C-to-A		
CC	transversion mutation sequence, and probes AAT17118 and AAT17120 bind		
CC	to the corresponding normal sequence. Primers 485 (AAT17122) and 502		
CC	(AAT17123) may be used along with primer 348 (AAT17121) to amplify		
CC	mutant and normal sequences, respectively, by PCR. Mutations in the		
CC	retinal degeneration slow protein and retinal rod		
CC	cGMP-phosphodiesterase genes are also implicated in retinitis		
CC	pigmentosa. Detection of any of these mutations in a fetus or		
CC	patient may be used in diagnosis.		
XX			
SQ	Sequence 3016 BP; 689 A; 863 C; 753 G; 711 T; 0 other:		
XX			
XX	Best Match	89.2%; Score 21.4; DB 17; Length 3016;	
XX	Query Local Similarity	95.7%; Pred. No. 1.7;	
XX	Matches 22; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
OY	1 GCCTTCTTTGCCAAGAGCGCGC 23		
DB	1168 GCGCTTCCTTGCCAAGAGCGCGC 1190		
XX			
XX	RESULT 2		
XX	AAT17116		
ID	AAT17116 standard; DNA: 6953 BP.		
XX			
AC	AAT17116;		
XX			
DF	06-JUL-1996 (first entry)		
XX			
DE	Rhodopsin gene.		
XX			

KW	Human; rhodopsin; transversion; mutation; retinitis pigmentosa;
KV	Intron; probe; primer hybridisation; polymerase chain reaction; PCR;
RN	eye; rod; retina; diagnostic; prenatal diagnosis; photoreceptor; ds.
XX	
OS	Homo sapiens.
XX	
PH	Location/Qualifiers
FT	Key
FT	5'UTR
FT	/**tag= a
FT	200..294
FT	5'UTR
FT	202..294
FT	/**tag= b
FT	/note= "Alternative 5'-UTR"
FT	complement (231..250)
FT	/**tag= c
FT	/note= "Binds primer 348 (AAT17121)"
FT	295..655
FT	/**tag= d
FT	/number= 1
FT	complement (354..372)
FT	/**tag= e
FT	/note= "Binds probes AAT17117 (mutant) and AAT17118 (normal)"
FT	complement (355..369)
FT	/**tag= f
FT	/note= "Binds probes AAT17119 (mutant) and AAT17120 (normal)"
FT	362
FT	/**tag= g
FT	/note= "Substitution with A in mutant sequence"
FT	362..381
FT	/**tag= h
FT	/note= "Binds primers 485 (AAT17122) (normal) and 502 (mutant)"
FT	666..2438
FT	/**tag= i
FT	/number= 1
FT	2439..2607
FT	/**tag= j
FT	/number= 2
FT	2608..3812
FT	/**tag= k
FT	/number= 2
FT	3813..3978
FT	/**tag= l
FT	/number= 3
FT	3979..4094
FT	/**tag= m
FT	/number= 3
FT	4095..4334
FT	/**tag= n
FT	/number= 4
FT	4335..5167
FT	/**tag= o
FT	/number= 4
FT	5168..5278
FT	/**tag= p
FT	/number= 5
PN	US5498521-A.
XX	
PD	12-MAR-1996.
PF	24-JAN-1990;
PE	90US-0469215.
XX	
PR	11-MAR-1993;
PR	93US-0033081.
PR	24-JAN-1990;
PR	90US-0469215.
PR	11-DEC-1991;
PR	91US-0805123.
PA	(HARD) HARVARD COLLEGE.
XX	
PI	Berson EL, Dryja TP;
XX	
DR	WPI; 1996-159684/16.

PT Diagnosis of hereditary retinal degenerative diseases e.g. retinitis pigmentosa, - caused by a human photoreceptor protein mutation, by detection of the mutation by PCR amplification or hybridisation methods

Example 1; Column 23-30; 71pp; English.

This sequence encodes human rhodopsin, and is shown in full with introns. The corresponding sequence without introns is shown in AAT17116. Substitution of histidine for the normal nonpolar amino acid proline at position 23, by substitution of C with A in codon 23, results in a dysfunctional or absent molecule, affecting rod function, and is linked with autosomal dominant retinitis pigmentosa. Probes AAT17117 and AAT17119 bind to the C-to-A transversion mutation sequence, and probes AAT17118 and AAT17120 bind to the corresponding normal sequence. Primers 485 (AAT17122) and 502 (AAT17123) may be used along with primer 348 (AAT17121) to amplify mutant and normal sequences, respectively, by PCR. Mutations in the retinal degeneration slow protein and retinal rod cGMP-phosphodiesterase genes are also implicated in retinitis pigmentosa. Detection of any of these mutations in a foetus or patient may be used in diagnosis.

sequence 6953 BP; 1523 A; 2022 C; 1797 G; 1611 T; 0 other;

Query Match	89.28;	Score 21.4;	DB 17;	Length 6953;
Best Local Similarity	95.78;	Pred. No. 2;		
Matches 22; Conservative	0.0	Mismatches 1		

	misassemblies	Indels	Gaps
1 GCATTCTTTGCCAAGGCGCCGC 23			
4272 GCGTTCCTTGCCAAAGGCGCCGC 4294			

RESULT 3

AAx84344
 AAX84344 standard; DNA; 8641 BP.
 X

AAx84344;

08-SEP-1999 (first entry)

stealth virus nucleic acid clone, SEQ ID NO: 36.

infection; diagnosis; infection; ss.

WO9934019-A1.

08-JUL-1999.

30-DEC-1998; 98WO-US27744.

9/05-0001184.

Mar 11 1977

WPI; 1999-405521/34.

Novel strains of stealth virus

Claim 19; Page 76-79; 95pp; English.

This sequence represents a Stealth virus nucleic acid clone. The invention relates to a method of detecting and characterising a stealth virus by reacting a sample suspected of containing a stealth virus with a probe from a known stealth virus and sequencing the resultant isolated nucleotide. The method comprises the steps of: (a) isolating DNA or RNA

CC from a sample suspected of containing a stealth virus, e.g. a culture of
CC cells showing a viral cytopathic effect; (b) testing the reactivity of
CC the isolated DNA or RNA with a molecular probe that contains at least 18
CC or more contiguous nucleotides identical to sequence previously
CC identified from a stealth virus; and, optionally (c) sequencing the
CC isolated DNA or RNA molecules that react with the probe. The method is
CC used to detect stealth virus in a biological product, food or in the
CC environment. The method is also used to evaluate agents for their
CC inhibitory or stimulatory effects on stealth virus replication and to
CC determine capacity of the virus to recombine with and potentially alter
CC the nucleic acid sequences of a cell or bacterium.

Sequence 8641 BP; 2101 A; 2031 C; 2018 G; 2476 T; 15 other;

Query Match	78.3%	Score 18.8;	DB 20;	Length 8641;
Best Local Similarity	90.9%	Pred. No. 37;		
Matches 20; Conservative	0;	Mismatches 2;	Totals 0;	

Oy	2	CTTTCCTTGGCAGAGCGCCG	23
Db	8149	CTTTCCTTGGCAGAGCGCCG	8170

RESULT 4
AAQ43543

ID	standard; CDNA; 3129 BP
AA043543	
XX	

AAQ43543;

11-NOV-1993 (first entry)

Rhodopsin gene.

Human; rhodopsin; mutant; retinal degeneration; primer; probe; hereditary; ss.

Homo sapiens.

key	Location/Qualifiers
prim_transcript	200..1341

```
/*tag = a
295..1341
```

WO9312134-A

24-JUN-1993

08-DEC-1992;

11-DEC-1991; 91US-0805123

(HARD) HARVARD COLLEGE.
Beyson Et. Davis an

WPT; 1993-214088/26.

P-PSDB; AAR38483.

probe or primer conty. sequence of human retinal degeneration slow protein mutant - used to diagnose hereditary retinal degenerative diseases

Disclosure; Fig 1; 56

The sequence given represents the human rhodopsin cDNA. Mutant versions of this sequence encode proteins which cause retinal degeneration. These sequences may be identified using 5' primers/probes described in the invention (see also AMO4545-9) and may be used to diagnose hereditary retinal degeneration. This sequence is the closest approximation to the gene sequence as the sequence is in the specification is not printed clearly.

XX
AC AAS41923;

```
XX 17-DEC-2001 (first entry)
DF
XX
XX Genomic sequence #239 encoding novel human enzyme polypeptide.
DE
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW lysase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotoxic; anticoagulant; ds.
XX
OS Homo sapiens.
XX
XX MO200153301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001MO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 15-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205115.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 11-JUL-2000; 2000US-0216880.
XX 14-JUL-2000; 2000US-0217487.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 14-AUG-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225265.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225272.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 18-AUG-2000; 2000US-0225759.
XX 22-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 23-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227189.
XX 30-AUG-2000; 2000US-0228294.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
PR
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0244676.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
```

[illegible]

PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225265.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225475.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0228287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0230458.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 25-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0234998.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 02-OCT-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 03-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 20-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0241826.
 PR 08-NOV-2000; 2000US-0246417.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249289.
 PR 17-NOV-2000; 2000US-0249289.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPL; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and
 metastasis -

Disclosure: SEQ ID NO 39030; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
 activity, and can be used in gene therapy and vaccine production. (I)
 proteins and polynucleotides may be used in the prevention, diagnosis and
 treatment of diseases associated with inappropriate (I) expression. For
 example, they may be used to treat disorders associated with decreased
 expression by rectifying mutations or deletions in a patient's genome
 that affect the activity of (I) by expressing inactive proteins or to
 supplement the patient's own production of (I). Additionally, (I)
 polynucleotides may be used to produce the secreted (I), by inserting
 the nucleic acids into a host cell and culturing the cell to express the
 protein. (I) proteins and polynucleotides may be used to prevent,
 diagnose and treat immune/hematopoietic-related diseases, especially
 cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 to AAK67694 represent human immune/hematopoietic antigen genomic
 sequences from the present invention. AAK54942 to AAK54950 and AAK62169
 represent sequences used in the exemplification of the present invention.

XX Sequence 36135 BP; 10321 A; 8075 C; 8063 G; 9676 T; 0 other;
 SQ
 Query Match 70.8%; Score 17; DB 22; Length 36135;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CTTCTTGGCAAGAGC 18
 |||||||
 DB 920 CTTCTTGGCAAGAGC 904

RESULT 10
 ABV12459
 ID ABV12459 standard; cDNA; 467 BP.
 XX
 AC ABV12459;
 XX
 DT 13-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker CDNA 12450.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI: 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 2053; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

SQ Sequence 467 BP; 71 A; 115 C; 132 G; 149 T; 0 other;
 Query Match 70.0%; Score 16.8; DB 23; Length 467;
 Best Local Similarity 90.0%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCTTCTTGGCAAGAGCGC 20
 |||||||
 DB 330 GCTTCTTGGCAAGAGCGC 349

RESULT 11
 ABV33599
 ID ABV33599 standard; cDNA; 506 BP.
 XX
 AC ABV33599;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker CDNA 33590.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI: 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 7105; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

SQ Sequence 506 BP; 80 A; 129 C; 145 G; 152 T; 0 other;
 Query Match 70.0%; Score 16.8; DB 23; Length 506;
 Best Local Similarity 90.0%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCTTCTTGGCAAGAGCGC 20
 |||||||
 DB 376 GCTTCTTGGCAAGAGCGC 395

RESULT 12
 ABK63732
 ID ABK63732 standard; cDNA; 1679 BP.
 XX
 AC ABK63732;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Rat sequence differentially expressed in response to a hepatotoxin #1639.
 XX
 KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 KM differential expression; centrilobular necrosis; steatosis.
 XX
 OS Rattus norvegicus.
 XX
 PN WO200210453-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 30-JUL-2001; 2001WO-US23872.
 XX
 PR 31-JUL-2000; 2000US-222040P.
 XX
 PR 02-NOV-2000; 2000US-244880P.
 PR 11-MAY-2001; 2001US-290022P.
 PR 15-MAY-2001; 2001US-290645P.
 PR 22-MAY-2001; 2001US-292336P.
 PR 06-JUN-2001; 2001US-295798P.
 PR 13-JUN-2001; 2001US-297457P.
 PR 19-JUN-2001; 2001US-298884P.
 PR 09-JUL-2001; 2001US-303455P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
 XX
 DR WPI; 2002-241625/29.
 XX
 PT Predicting toxic effects of compounds or the progression of these toxic
 PT effects by determining the changes in gene expression in tissues or
 PT cells exposed to the toxin and comparing these to gene expression in
 PT unexposed tissues or cells -
 XX
 PS Claim 1; Seq ID No 1639; 239pp; English.
 XX
 CC The invention relates to methods for predicting toxic effects of
 CC compounds or the progression of these toxic effects by determining the
 CC global changes in gene expression in tissues or cells exposed to the
 CC toxin and comparing these to gene expression in unexposed tissues or
 CC cells. Also included are methods of predicting at least one toxic
 CC effect of a compound or progression of a toxic effect, preferably the
 CC hepatotoxicity of a compound, comprising detecting the level of
 CC expression in a tissue or cell sample exposed to the compound of two or
 CC more genes listed in the specification, where differential expression of
 CC the genes is indicative of at least one toxic effect or progression.
 CC The method can also be used to identify an agent which modulates the
 CC toxic response and predict cellular pathways that a compound modulates
 CC in a cell. The methods utilize a set of at least two probes (on a solid
 CC support in kit form), where each of the probes comprises a sequence that
 CC specifically hybridizes to a gene listed in the specification, a computer
 CC system comprising a database containing information identifying the
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
 CC set of genes comprising at least two genes listed in the specification,
 CC and a user interface to view the information used to present information,
 CC identifying the expression level in a tissue or cell of at least one gene
 CC listed in the specification. The method is useful for elucidating global
 CC changes in gene expression and for identifying toxicly markers in
 CC tissues or cell exposed to a known toxin. The genes may be used as
 CC toxicity markers in drug screening and toxicity assays. The genes and
 CC gene expression information may be used as diagnostic markers for the
 CC prediction or identification of the physiological state of tissue or cell
 CC sample that has been exposed to a compound or agent. Hepatotoxicity
 CC is characterized by centrilobular necrosis and steatosis. The present
 CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene

CC which is differentially expressed in response to a hepatotoxic agent.
 XX
 SQ Sequence 1679 BP; 377 A; 471 C; 483 G; 348 T; 0 other;

Query Match 70.0%; Score 16.8; DB 24; Length 1679;
 Best Local Similarity 90.0%; Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 TCTTGGCCAGAGCGCCGCA 24
 ||||||| |||||||||
 Db 61 TCTTGGCCAGAGCGCCGCA 80

RESULT 13
 ABL02991
 ID ABL02991 standard; cDNA; 2304 BP.
 XX
 AC ABL02991;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 3455.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09221.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 DR P-PSDB; ABB58888.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 3455; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57727-AB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX

SQ Sequence 2304 BP; 541 A; 640 C; 721 G; 402 T; 0 other;

Query Match 70.0%; Score 16.8; DB 23; Length 2304;
 Best Local Similarity 90.0%; Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TCTTGGCCAGAGCGCCGCGC 23
 ||||||| |||||||||
 Db 226 TCTTGGCCAGAGCGCGCAGC 245

RESULT 14
ABL02990
ID ABL02990 standard; cDNA; 4446 BP.
XX
XX ABL02990;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 3452.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
XX
XX pharmaceutical; gene; ss.
OS
XX
XX Drosophila melanogaster.
PN
XX
XX WO200171042-A2.
PD
XX
XX 27-SEP-2001.
PF
XX
XX 23-MAR-2001; 2001WO-US09231.
PR
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI: 2001-656860/75.
DR
XX
XX P-PSDB; ABB58887.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
XX Claim 1: SEQ ID NO 3452; 21bp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
CC sequences (AB101840-ABL16175) and the encoded proteins
CC (AB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
XX Sequence 4446 BP; 1191 A; 1023 C; 1084 G; 1148 T; 0 other;
SQ
XX
XX Query Match 70.0%; Score 16.8; DB 23; Length 4446;
Best Local Similarity 90.0%; Pred. No. 3e+02; 2; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 2;
OY 4 TTCTTCCCAAGAGCGCCG 23
DB 613 TTCTTCCCAAGAGCGCAGC 632

RESULT 15
ABK80729
ID ABK80729 standard; DNA; 408 BP.
XX
XX ABR80729;
AC
XX
XX 13-AUG-2002 (first entry)
DT
XX
XX Bacillus clausii genomic sequence tag (GST) #3572.
DE
XX
XX Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
XX
XX Bacillus clausii.
OS
XX
XX WO200229113-A2.
PN
XX
XX 11-APR-2002.
PD
XX
XX 05-OCT-2001; 2001WO-US31437.
PF
XX
XX 06-OCT-2000; 2000US-0680598.
PR
XX
XX 27-MAR-2001; 2001US-279526P.
PR
XX
XX (NOVO) NOVOZYMES BIOTECH INC.
PA
XX
XX (NOVO) NOVOZYMES AS.
PA
XX
XX Bertka R, Clausen IG;
PI
XX
XX WPI: 2002-416684/44.
DR
XX
XX
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -
XX
XX
XX Claim 11: SEQ ID NO 8020; 200pp; English.
PS
XX
XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions.
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
XX Sequence 408 BP; 106 A; 86 C; 104 G; 111 T; 1 other;
SQ
XX
XX Query Match 69.2%; Score 16.6; DB 24; Length 408;
Best Local Similarity 82.6%; Pred. No. 2.5e+02; 4; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 4;
OY 1 GCTTCTTCCCAAGAGCGCCG 23
DB 374 GCATTCTTCCCAAGTGCAGC 396

Search completed: March 17, 2003, 10:50:45
Job time : 146.715 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 848.387 seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-3
Perfect score: 24
Sequence: 1 gcttccttcgcaagagcgccgca 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21.4	89.2	303	14	BM694379 UI-E-CII-
2	21.4	89.2	319	14	BM694379 UI-E-CII-
3	21.4	89.2	337	14	BM694379 UI-E-CII-
4	21.4	89.2	340	14	BM694379 UI-E-CII-
5	21.4	89.2	349	14	BM694379 UI-E-CII-
6	21.4	89.2	441	14	BM694379 UI-E-CII-

Result 1	Score	Query Match	Length	DB ID	Description
7	21.4	89.2	442	9	AL712412 DKEP6860
8	21.4	89.2	446	14	BM696193 UI-E-CII-
9	21.4	89.2	469	14	BM688227 UI-E-CII-
10	21.4	89.2	471	9	AL712402 DKEP686N
11	21.4	89.2	471	14	BM688069 UI-E-CII-
12	21.4	89.2	474	14	BM688069 UI-E-CII-
13	21.4	89.2	478	13	BM662914 UI-E-CII-
14	21.4	89.2	493	14	BM690137 UI-E-CII-
15	21.4	89.2	493	14	BM691592 UI-E-CII-
16	21.4	89.2	496	14	BM690311 UI-E-CII-
17	21.4	89.2	499	14	BM690311 UI-E-CII-
18	21.4	89.2	503	14	BM690311 UI-E-CII-
19	21.4	89.2	504	14	BM690311 UI-E-CII-
20	21.4	89.2	505	14	BM690311 UI-E-CII-
21	21.4	89.2	510	14	BM690311 UI-E-CII-
22	21.4	89.2	513	14	BM690311 UI-E-CII-
23	21.4	89.2	514	14	BM690311 UI-E-CII-
24	21.4	89.2	516	14	BM690311 UI-E-CII-
25	21.4	89.2	526	14	BM690311 UI-E-CII-
26	21.4	89.2	530	14	BM690311 UI-E-CII-
27	21.4	89.2	530	14	BM690311 UI-E-CII-
28	21.4	89.2	535	14	BM690311 UI-E-CII-
29	21.4	89.2	549	14	BM690311 UI-E-CII-
30	21.4	89.2	549	14	BM690311 UI-E-CII-
31	21.4	89.2	562	14	BM690311 UI-E-CII-
32	21.4	89.2	575	14	BM690311 UI-E-CII-
33	21.4	89.2	576	14	BM690311 UI-E-CII-
34	21.4	89.2	578	14	BM690311 UI-E-CII-
35	21.4	89.2	580	14	BM690311 UI-E-CII-
36	21.4	89.2	580	14	BM690311 UI-E-CII-
37	21.4	89.2	587	14	BM690311 UI-E-CII-
38	21.4	89.2	590	14	BM690311 UI-E-CII-
39	21.4	89.2	591	9	AL711507 DKEP686D
40	21.4	89.2	595	14	BM690311 UI-E-CII-
41	21.4	89.2	599	14	BM690311 UI-E-CII-
42	21.4	89.2	604	14	BM690311 UI-E-CII-
43	21.4	89.2	605	9	AL712251 DKEP686L
44	21.4	89.2	606	14	BM690311 UI-E-CII-
45	21.4	89.2	606	14	BM690311 UI-E-CII-

ALIGNMENTS

RESULT 1
LOCUS BM694379 303 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-CII-afp-p-18-0-UI-r1 UI-E-CII Homo sapiens cDNA clone
ACCESSION BM694379
VERSION 1
KEYWORDS BM694379.1 GI:19007637
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 303)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. W. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

Seq primer: M13 Reverse.

FEATURES

Source

Location/Qualifiers

1..303

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-C11-afp-p-18-0-UI"

/clone_lib="UI-E-C11"

/tissue_type="RPE and Choroid"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site: 1: Ecor I; Site: 2: Not I;

UI-E-C11 is a normalized cDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag

sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACCCTA.

This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT
ORIGIN
70 a 101 c 67 g 65 t

Query Match 89.2%; Score 21.4; DB 14; Length 303;
Best Local Similarity 95.7%; Pred. No. 23;

Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGCCAGAGCGCCG 23

Db 171 GCGTCTTGCCAGAGCGCCG 193

RESULT 2

LOCUS

B0637565 319 bp mRNA linear EST 15-JUL-2002

DEFINITION hel1904.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he

Accession B0637565

VERSION B0637565.1 GI:21762024

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 11 row: 9 column: 04

Seq primer: M13RPI reverse primer (ABI).

Location/Qualifiers

1..319

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="hel1904"

/clone_lib="Human Retina cDNA (Un-normalized, unamplified)

/tissue_type="Retina"

/dev_stage="Adult"

/lab_host="EMDH10B"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed

eye disease. 100ug of total RNA was used for library

construction. A directionally cloned cDNA library in the

pSPORT1 vector (Life Technologies) was constructed at

Bioserve Biotechnology (Laurel MD) essentially following

the protocols of the SuperScript Plasmid System full

details of which are contained in the manufacturer's

Instruction manual (http://www.lifetech.com/). First

strand synthesis was carried out using a Not I

primer-adaptor [5'-TGACGTGTTGATGTCAGCGCCG(7)15-3'

]. EST analysis was performed on the unamplified library

at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT 72 a 108 c 75 g 64 t

ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 319;

Best Local Similarity 95.7%; Pred. No. 23;

Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGCCAGAGCGCCG 23

Db 204 GCGTCTTGCCAGAGCGCCG 226

RESULT 3

LOCUS BM723222 337 bp mRNA linear EST 01-MAR-2002

DEFINITION UI-E-EJ0-1-15-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone

UI-E-EJ0-1-15-0-UI 5', mRNA sequence.

Accession BM723222

VERSION BM723222.1 GI:19044349

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CONTACT: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA

sequence: 295-332, >AT-richLowComplexity (matched complement)

Seq primer: M13 Reverse.

Location/Qualifiers

1..337

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJ0-1-15-0-UI"

/clone_lib="UI-E-EJ0"

/tissue_type="fetal eyes, lens, eye anterior segment,

optic nerve, retina, Retina foveal and Macular, RPE and

Choroid"

	Query Match	89.2%	Score 21.4	DB 14	Length 337
	Best Local Similarity	95.7%	Pred. NO. 24		
	Matches 22	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1 GCCTTCTTTGCCAAGAGCGCCGC	23			
Db	1 GCCTTCTTTGCCAAGAGCGCCGC	23			

RESULT 4	
BM682444/c	
LOCUS	BM682444
DEFINITION	340 bp mRNA linear EST 27-FEB-2002
UI-E-EUO-a10-1-15-0-UI s1	UI-E-EUO Homo sapiens CDNA clone
UI-E-EUO-a10-1-15-0-UI 3'	mRNA sequence.
BM682444	
ACCESSION	BM682444

VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT

BM682444.1 GI:18992340
 EST.
 human.
 Homo sapiens
 Eumetazoa: Chordata: Craniata: Vertebrata: Euteleostomi
 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
 1 (bases 1 to 340)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB

JOURNAL
MEDLINE
COMMENT

Genome Res. 6 (9), 791-806 (1996)
97044477

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

Tissue procurement: Dr. Gregg Hagaman
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library arrayed by: Dr. M. Bento Soares, University of Iowa
DNA sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).

The following repetitive elements were found in this cDNA
sequence: 1-37, >AT-rich#low-complexity (matched complement)
Seq primer: M13 Forward
POLYA=yes.

FEATURES
source

```
Location/Qualifiers
1..340
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-alo-1.15-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment
```

```

BASE COUNT      71 a      76 c      105 g      88 t
ORIGIN
      optic nerve, retina, Retina Foveal and Macular, RPE and
      Choroid"
      /dev_host="fetal and adult"
      /lab_host="PH10B (Life Technologies) (T1 phage resistant)"
      /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
      modified polylinker; Site.1: Ecor I; Site.2: Not I;
      ui-E-EJ0 is a subcloned cDNA library constructed
      according to Bonaldo, Lennon and Soares, Genome Research,
      6,791-806, 1996. First strand cDNA synthesis was primed
      with an oligo-dT primer containing a Not I site. Double
      stranded cDNA was ligated to an Ecor I adaptor, digested
      with Not I, and cloned directionally into pT73-Pac
      vector. The oligonucleotide used to prime the synthesis of
      first-strand cDNA contains a library tag sequence that is
      located between the Not I site and the (dT)18 tail. The
      sequence tags for this library are: fetal eyes, AGATTACAGC
      ; lens, CGATTAGCCG; eye anterior segment, ATGCGGCAT;
      optic nerve, CCATTAGGTG; retina, CCGCG; Retina Foveal and
      Macular, GTCC; RPE and Choroid, ACCCA. This library was
      created for the program, Gene Discovery in the Visual
      System, supported by National Eye Institute (NEI).
      TAG LIB-UI-E-EJ0
      TAG TISSUE-human retina
      TAG_SEQ-CCGCG"

```

	Query Match	89.28%	Score 21.4	DB 14	Length 340
	Best Local Similarity	95.7%	Pred. No. 24		
	Matches 22	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1 GCCTTCCTTCCCAAGAGCGCGC	23			
DB	331 GCGTCTCTTCCCAAGAGCGCGC	309			

RESULT 5	BM703950	349 bp	RNA	linear	EST 28-FEB-2002
LOCUS	BM703950				
DEFINITION	BM703950	349 bp	RNA	linear	EST 28-FEB-2002
ACCESSION	UI-E-CK1-atk-m-09-0-UI.r1	UI-E-CK1	Homo sapiens	CDNA clone	
VERSION	BM703950				
KEYWORDS	BM703950.1	GI:19017208			
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	1 (bases 1 to 349)				
JOURNAL	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
MEDLINE	Normalization and subtraction: two approaches to facilitate gene				
COMMENT	discovery				
	Genome Res. 6 (3), 791-806 (1996)				
	9704477				
	Contact: Soares, MB				

COMMENT

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES
SOURCE

```
Location/Qualifiers
1. 349
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CK1-afk-m-09-0-UI
```

	Query Match	89.2%	Score 21.4	DB 14	Length 349
	Best Local Similarity	95.7%	Pred. No. 24		
	Matches 22	Conservative 0	Mismatches 1	Indels 0	Gaps 0
OY	1 GCTTCTTGGCCAGAGCGCCGC	23			
Db	37 GCGTCTTTGGCCAGAGCGCCGC	59			

ACCESSION	BM690151	GI:19003409
VERSION	BM690151.1	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE	1 (bases 1 to 441)
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Soares, MB

451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscores@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Greg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
1. .441

```

FEATURES
source
location/Qualifiers
1..441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="vii-E-ClO-acid-f-11-0-UI"
/clone_lib="vii-E-ClO"
/tissue_type="human retina"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

```

Dy 1 GCTTCTTGCCAGAGCGCCG 23
|| |||||||
Db 301 GCGTCTTGCCAGAGCGCCG 322

RESULT 7	AL712412	442 bp	mRNA	linear	EST 22-MAR-2002
LOCUS	AL712412				
DEFINITION	DKFZdp66601888.r1.686 (synonym: h1ccc3) Homo sapiens cDNA clone				
ACCESSION	AL712412				
VERSION	AL712412.1				
KEYWORDS	EST.				
SOURCE	human.				

ORGANISM *Homo sapiens*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 442)
AUTHORS Wambutt,R., Heubner,D., Mewes,W., Well,B. and Wiemann,S.
TITLE EST (Wambutt,R., Heubner,D., Mewes,H.W., Well,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT
MIPS

FEATURES

Am Klopferstr. 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by AGOMA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp6601888) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers

```

SOURCE
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686O1888"
/clone_1b="686 (synonym: hicc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Vector: pT7A1pEx2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
BASE COUNT      106 a      144 c      97 g      95 t
ORIGIN

```

Query Match	89.2%	score	21.4	DB	9	length	442
Best Local Similarity	95.7%	Pred. No.	26				
Matches	22	Conservative	0	Mismatches	1	Indels	0
				Gaps			0

OY 1 GCTTCTTTGCCAAGAGCGCCG 23
 Db 23 GCCTTCTTTGCCAAGAGCGCCG 45

RESULT 8 BM696193 LOCUS

DEFINITION BM696193 446 bp mRNA linear EST 28-FEB-2002
 UI-E-CL1-afa-d-24-0-UI r1 UI-E-CL1 Homo sapiens cDNA clone

ACCESSION BM696193
 UI-E-CL1-afa-d-24-0-UI 5', mRNA sequence.

VERSION BM696193.1 GI:19009451
 EST.

KEYWORDS
 SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 446)
 TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477

COMMENT

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES

Source

Location/Qualifiers

1..446

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-CL1-afa-d-24-0-UI"

/clone_lib="UI-E-CL1"

/tissue_type="human retina"

/dev_stage="adult"

/lab_host="PH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-CL1 is a normalized cDNA library containing the
 following tissue(s): retina. The library was constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CCGCG. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

BASE COUNT

ORIGIN 95 a 148 c 106 g 96 t 1 others

Query Match

Best Local Similarity 89.2%; Score 21.4; DB 14; Length 446;
 Pred. No. 26;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCTTCTTTGCCAAGAGCGCCG 23
 Db 23 GCCTTCTTTGCCAAGAGCGCCG 248

RESULT 9 BM688227 LOCUS

DEFINITION BM688227 469 bp mRNA linear EST 28-FEB-2002
 UI-E-CL0-aby-g-03-0-UI r1 UI-E-CL0 Homo sapiens cDNA clone

ACCESSION BM688227
 UI-E-CL0-aby-g-03-0-UI 5', mRNA sequence.

VERSION BM688227.1 GI:19001478
 EST.

KEYWORDS
 SOURCE

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 469)
 TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477

COMMENT

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this CDNA
 Sequence: 425-459, >AT-rich#low_complexity
 Seq primer: M13 Reverse.

FEATURES

Source

Location/Qualifiers

1..469

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-CL0-aby-g-03-0-UI"

/clone_lib="UI-E-CL0"

/tissue_type="human retina"

/dev_stage="adult"

/lab_host="PH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-CL0 is a cDNA library containing the following
 tissue(s): retina. The library was constructed according
 to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT73-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CCGCG. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

BASE COUNT

ORIGIN 110 a 155 c 99 g 104 t 1 others

Query Match

Best Local Similarity 89.2%; Score 21.4; DB 14; Length 469;
 Pred. No. 27;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCTTCTTTGCCAAGAGCGCCG 23
 Db 82 GCCTTCTTTGCCAAGAGCGCCG 104

RESULT 10 AL712402 LOCUS

DEFINITION AL712402 471 bp mRNA linear EST 22-MAR-2002

DEFINITION DKFZp686N1788.r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
 ACCESSION AL712402
 VERSION AL712402.1 GI:19695757
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 471)
 Mambutt, R., Heubner, D., Mewes, W., Well, B. and Wiemann, S.
 EST (Mambutt, R., Heubner, D., Mewes, H.W., Well, B. and Wiemann, S.)
 Unpublished (1999)
 CONTACT: Mambutt R
 MIPs
 Am Klopfersplitz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by AGOMA (Berlin/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.
 This clone (DKFZp686N1788) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers
 1..471
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp686N1788"
 /clone_lib="686 (synonym: hlcc3)"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
 cDNA-collection"
 CDNA-collection"
 BASE COUNT 109 a 158 c 99 g 105 t
 ORIGIN
 Query Match 89.2%; Score 21.4; DB 9; Length 471;
 Best Local Similarity 95.7%; Pred. No. 27;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GCTTCTTGGCAGAGCGCCG 23
 11 |||||
 Db 23 GCGTCTTGGCAGAGCGCCG 45
 RESULT 11 471 bp mRNA linear EST 03-MAY-2002
 B0250368
 LOCUS TAE25006A11R TAE25 Triticum aestivum cDNA clone TAE25006A11R, mRNA
 DEFINITION
 ACCESSION B0250368
 VERSION B0250368.1 GI:20446244
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 ; Triticeae; Triticum.
 1 (bases 1 to 471)
 Cloutier, S.
 Wheat functional genomics - Glenlea developing seeds cDNA libraries
 Unpublished (2002)
 CONTACT: Dr. Sylvie Cloutier
 Cereal Research Centre, Agriculture and Agri-food Canada
 195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M5
 Tel: (204) 983-2340
 Fax: (204) 983-4604
 Email: scloutier@em.agr.ca
 was cloned directionally, not all sequences generated with reverse
 primer were from the 5' end (same with forward primer and 3' end).

Average insert size is >870 bp
 Plate: 006 row: A column: 11
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..471
 /organism="Triticum aestivum"
 /cultivar="Glenlea"
 /db_xref="taxon:4565"
 /clone="TAE25006A11R"
 /clone_lib="TAE25"
 /tissue_type="developing seeds"
 /dev_stage="25 days after anthesis"
 /lab_host="E. coli DH10B"
 /note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
 Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
 of cultivar Glenlea 25 days post-anthesis"
 BASE COUNT 108 a 164 c 108 g 91 t
 ORIGIN
 Query Match 89.2%; Score 21.4; DB 14; Length 471;
 Best Local Similarity 95.7%; Pred. No. 27;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GCTTCTTGGCAGAGCGCCG 23
 11 |||||
 Db 174 GCGTCTTGGCAGAGCGCCG 196
 RESULT 12 474 bp mRNA linear EST 28-FEB-2002
 B0688069
 LOCUS UI-E-C10-abv-c-04-0-UI.r1 UI-E-C10 Homo sapiens cDNA clone
 DEFINITION
 ACCESSION B0688069
 VERSION UI-E-C10-abv-c-04-0-UI 5', mRNA sequence.
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 474)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL MEDLINE
 CONTACT: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msquares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 335-367, >AT-richLow-complexity
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..474
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-C10-abv-c-04-0-UI"
 /clone_lib="UI-E-C10"
 /tissue_type="human retina"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: p773-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;

Location/Qualifiers
1..478
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-Ck0-aan-e-04-0-0-ui"
/clone_11b="UI-E-Ck0"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ eye: Vector: pRTT3-Pac (pharmacia) with a modified polylinker. Site_1: EcoR I; Site_2: Not I; UI-E-Ck0 is a cDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA

```

Location/Qualifiers
1. 493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CLO-acd-d-08-0-UI"
/clone_lib="UI-E-CLO"
/tissue_type="human retina"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pRTT3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-CLO is a cDNA library containing the following
tissue(s): retina. The library was constructed according
to Bonaldo, Lennon and Soares, Genome Research, 6:791-806

```

1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CCGCG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 110 a 162 c 108 g 113 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 493;
Best Local Similarity 95.7%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCTTCTTTGCCAAGAGCGCGC 23
DB 131 GCGTCTTTGCCAAGAGCGCGC 153

RESULT 15 493 bp mRNA linear EST 28-FEB-2002
BM691592
LOCUS UI-E-C11-abh-c-07-0-UI.r1 UI-E-C11 Homo sapiens cDNA clone
DEFINITION UI-E-C11-abh-c-07-0-UI 5', mRNA sequence.
ACCESSION BM691592
VERSION BM691592.1 GI:19004850
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 493)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 302-335, >AT-rich#Low_complexity
Seq primer: M13 Reverse.

FEATURES
Source Location/Qualifiers

1..493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-C11-abh-c-07-0-UI"
/clone_lib="UI-E-C11"
/tissue_type="RPE and Choroid"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: p7773-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-C11 is a normalized cDNA library containing the following tissue(s): RPE and Choroid. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I

adaptor, digested with Not I, and cloned directionally into p7773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is ACCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 116 a 157 c 111 g 109 t
ORIGIN

Query Match 89.2%; Score 21.4; DB 14; Length 493;
Best Local Similarity 95.7%; Pred. NO. 27;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCTTCTTTGCCAAGAGCGCGC 23
DB 8 GCGTCTTTGCCAAGAGCGCGC 30

Search completed: March 17, 2003, 13:09:15
Job time : 851.387 secs

GenCore version 5.1.4-P5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 202.495 Seconds

(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-4

Sequence: 1 aagaaatacagacaagca 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 205640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1:  gb-ba:*
2:  gb-htg:*
3:  gb-in:*
4:  gb-om:*
5:  gb-ov:*
6:  gb-pat:*
7:  gb-ph:*
8:  gb-pl:*
9:  gb-pr:*
10: gb-ro:*
11: gb-sts:*
12: gb-sy:*
13: gb-un:*
14: gb-vi:*
15: em-ba:*
16: em-fun:*
17: em-hum:*
18: em-in:*
19: em-mu:*
20: em-om:*
21: em-or:*
22: em-ov:*
23: em-pat:*
24: em-ph:*
25: em-pl:*
26: em-ro:*
27: em-sts:*
28: em-un:*
29: em-vi:*
30: em-htg-hum:*
31: em-htg-inv:*
32: em-htg-mus:*
33: em-htg-other:*
34: em-htg-pin:*
35: em-htg-rod:*
36: em-htg-mam:*
37: em-htg-vrt:*
38: em-sy:*
39: em-hgo-hum:*
40: em-hgo-mus:*
41: em-hgo-other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	20.4	92.7	133095	2	AC046146	AC046146 Mus muscu
2	20.4	92.7	152147	2	AC095531	AC095531 Rattus no
3	20.4	92.7	172497	2	AC117913	AC117913 Rattus no
4	20.4	92.7	188795	2	AC113320	AC113320 Mus muscu
5	19.4	88.2	179767	9	AC069223	AC069223 Homo sapi
6	19.4	88.2	180638	9	AC068763	AC068763 Homo sapi
7	19.4	88.2	185841	2	AC027080	AC027080 Homo sapi
8	19.4	88.2	201416	2	AC130437	AC130437 Homo sapi
9	19.4	88.2	218859	2	AC023912	AC023912 Homo sapi
10	19.4	88.2	274349	2	AC093623	AC093623 Homo sapi
11	19.4	86.4	113940	2	AP005572	AP005572 Oryza sat
12	18.8	85.5	2052	5	AF012746	AF012746 Danio rer
13	18.8	85.5	2209	8	SCYPL252C	SCYPL252C Oryza sat
14	18.8	85.5	2411	8	SCYPL253C	SCYPL253C Oryza sat
15	18.8	85.5	37808	8	SC38CXV1	SC38CXV1 S. cerevisia
16	18.8	85.5	41578	3	CBR646G14	CBR646G14 Caenorhab
17	18.8	85.5	52495	2	AC109155	AC109155 Mus muscu
18	18.8	85.5	52912	2	AC130512	AC130512 Rattus no
19	18.8	85.5	63484	2	AC108916	AC108916 Mus muscu
20	18.8	85.5	64417	2	AC131268	AC131268 Homo sapi
21	18.8	85.5	96036	2	AC119622	AC119622 Homo sapi
22	18.8	85.5	97454	9	AL138147	AL138147 Rattus no
23	18.8	85.5	99374	9	AC092031	AC092031 Homo sapi
24	18.8	85.5	112632	2	AL137158	AL137158 Homo sapi
25	18.8	85.5	113955	9	AC068291	AC068291 Homo sapi
26	18.8	85.5	114457	2	AC095684	AC095684 Rattus no
27	18.8	85.5	120527	2	AP005487	AP005487 Oryza sat
28	18.8	85.5	132072	9	AC004841	AC004841 Homo sapi
29	18.8	85.5	140658	5	AF112374	AF112374 Danio rer
30	18.8	85.5	141138	2	AL845428	AL845428 Homo sapi
31	18.8	85.5	143496	2	AC111035	AC111035 Mus muscu
32	18.8	85.5	143823	9	AC019195	AC019195 Homo sapi
33	18.8	85.5	151040	2	AC068389	AC068389 Homo sapi
34	18.8	85.5	152161	9	AL359074	AL359074 Homo sapi
35	18.8	85.5	154348	2	AC069480	AC069480 Homo sapi
36	18.8	85.5	156195	9	AC093799	AC093799 Homo sapi
37	18.8	85.5	157038	2	AC099275	AC099275 Rattus no
38	18.8	85.5	157823	2	AL161795	AL161795 Homo sapi
39	18.8	85.5	158075	2	AC123076	AC123076 Rattus no
40	18.8	85.5	159386	2	AC098370	AC098370 Rattus no
41	18.8	85.5	163830	2	AC068235	AC068235 Homo sapi
42	18.8	85.5	166292	2	CNS080C8S	AL731888 Oryza sat
43	18.8	85.5	16764	2	AC115875	AC115875 Mus muscu
44	18.8	85.5	168199	9	CNS01DTS	AL132857 Human chr
45	18.8	85.5	169295	9	AL732602	AL732602 Human DNA

ALIGNMENTS

```

RESULT 1
AC046146 133095 bp DNA 1linear HTG 16-OCT-2001
LOCUS AC046146 Mus musculus chromosome 12 clone RP23-321N2L, *** SEQUENCING IN
DEFINITION AC046146
ACCESSION AC046146
VERSION AC046146.6 GI:16118085
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
1 (bases 1 to 133095)
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Bahay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, J., Forcum-Tansey, J., Gill, R.,
Gorrell, J. H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,
Hosack, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, D., Liu, W., Louised, H., Lozano, R. J., Martin, R.,
Massey, E., McLeod, M. P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Wallington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D. M., Rives, M., Scherer, S., Sodergren, E., Weinstein, G.,
Worley, K., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 133095)
Direct Submission
Submitted (13-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 14, 2001 this sequence version replaced gi:11094634.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center Project name: MADY
Center clone name: RP23-321N21
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye 3.1; 94% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 118378 bases at least Q40
Consensus quality: 144876 bases at least Q30
Consensus quality: 155160 bases at least Q20
Estimated insert size: 148882; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; average-of-estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 10336: contig of 10336 bp in length
* 10337 10436: gap of unknown length
* 10437 18208: contig of 8772 bp in length
* 18208 19308: gap of unknown length
* 19309 26724: contig of 7416 bp in length
* 26725 26824: gap of unknown length
* 26825 33512: contig of 6688 bp in length
* 33513 33612: gap of unknown length
* 33613 33688: contig of 5776 bp in length
* 33689 39488: gap of unknown length
* 39489 45330: contig of 5842 bp in length
* 45331 50898: gap of unknown length
* 50899 50998: gap of unknown length
* 50999 56466: gap of unknown length
* 56467 62096: contig of 5530 bp in length
* 62097 62196: gap of unknown length
* 62197 65971: contig of 3775 bp in length
* 65972 66071: gap of unknown length
* 66072 70109: contig of 4038 bp in length

70110 70209: gap of unknown length
* 70210 72595: contig of 2286 bp in length
* 72596 74792: gap of unknown length
* 74793 75072: contig of 2277 bp in length
* 75073 79255: gap of unknown length
* 79256 79355: contig of 4183 bp in length
* 79356 83184: contig of 3828 bp in length
* 83185 83283: gap of unknown length
* 83284 86781: contig of 3498 bp in length
* 86782 90021: contig of 3140 bp in length
* 90022 90121: gap of unknown length
* 90122 93597: contig of 3476 bp in length
* 93598 93698: gap of unknown length
* 93699 97229: contig of 3552 bp in length
* 97230 97329: gap of unknown length
* 97330 100516: contig of 3187 bp in length
* 100517 100616: gap of unknown length
* 100617 104190: contig of 3574 bp in length
* 104191 104290: gap of unknown length
* 104291 107955: contig of 3665 bp in length
* 107956 108055: gap of unknown length
* 108056 111930: contig of 3875 bp in length
* 111931 112030: gap of unknown length
* 112031 114125: contig of 2095 bp in length
* 114126 114225: gap of unknown length
* 114226 116520: contig of 2295 bp in length
* 116521 116620: gap of unknown length
* 116621 118829: contig of 2209 bp in length
* 118830 118930: gap of unknown length
* 118931 121281: contig of 2352 bp in length
* 121282 121381: gap of unknown length
* 121382 123762: contig of 2381 bp in length
* 123763 123862: gap of unknown length
* 123863 126537: contig of 2674 bp in length
* 126537 126636: gap of unknown length
* 126637 130483: contig of 3847 bp in length
* 130484 130583: gap of unknown length
* 130584 130584: contig of 2512 bp in length.

* Location/Qualifiers
* 1. 133095
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /chromosome="12"
* /clone="RP23-321N21"

BASE COUNT 34946 a 30809 c 29300 g 35021 t 3019 others
ORIGIN
Query Match 92.7%; Score 20.4; DB 2; Length 133095;
Best Local Similarity 95.5%; Pred. No. 65;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
*
* 1 AACGAAATCTAGACAAACAA 22
* |||||
* Db 71615 AACGCAATCTAGACAAACAA 71636

RESULT 2
AC095531
LOCUS
DEFINITION Rattus norvegicus clone CH230-8N2, *** SEQUENCING IN PROGRESS ***
AC095531
VERSION AC095531.3 GI:21716992
ACCESSION AC095531.3
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 152147)
REFERENCE
Muzny, D. M., Adams, C., Adio-Oduola, B., Ali-osman, F. R., Allen, C.,

Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbara, J., Benton, J., Blinage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowe, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Cartron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Den, A.L., Ding, Y., Dinh, H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Garrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homel, E., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogub, M., Okunolu, G.,
 Orgunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pichens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Syarik, A., Tabor, P., Tameris, A., Tameris, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 152147)
 Worley, K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 152147)
 Worley, K.C.
 Direct Submission
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 9, 2002 this sequence version replaced gi:17942049.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GCPY
 Center clone name: CH230-BN2
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 92864 bases at least Q40
 Consensus quality: 97631 bases at least Q30
 Consensus quality: 101839 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_dir/c_data.html).
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 58 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1151:	contig of 1151 bp in length
*	1152	gap of unknown length
*	1252	contig of 1454 bp in length
*	2705	gap of unknown length
*	2706	gap of unknown length
*	2806	contig of 1516 bp in length
*	4421:	gap of unknown length
*	4322	gap of unknown length
*	4422	contig of 1438 bp in length
*	5859:	gap of unknown length
*	5860	gap of unknown length
*	5860	contig of 1267 bp in length
*	7227	gap of unknown length
*	7327	gap of unknown length
*	8443:	contig of 1117 bp in length
*	8444	gap of unknown length
*	8544	contig of 1160 bp in length
*	9704	gap of unknown length
*	9804	contig of 1232 bp in length
*	11036	gap of unknown length
*	11136	contig of 1788 bp in length
*	11936	gap of unknown length
*	12924	contig of 1152 bp in length
*	13024	contig of 1152 bp in length
*	14176	gap of unknown length
*	14276	contig of 1937 bp in length
*	16212:	gap of unknown length
*	16213	gap of unknown length
*	16313	contig of 1579 bp in length
*	17892	gap of unknown length
*	17991:	contig of 1341 bp in length
*	19333	gap of unknown length
*	19433	contig of 1413 bp in length
*	20845:	gap of unknown length
*	20846	gap of unknown length
*	20945:	contig of 1838 bp in length
*	22096	gap of unknown length
*	22784	contig of 1036 bp in length
*	22884	contig of 1036 bp in length
*	23919:	gap of unknown length
*	24019:	contig of 1169 bp in length
*	24020	contig of 1169 bp in length
*	25189	gap of unknown length
*	25289	contig of 1146 bp in length
*	26435:	gap of unknown length
*	26535	contig of 1405 bp in length
*	27939:	gap of unknown length
*	27940	contig of 1781 bp in length
*	28040	contig of 1781 bp in length
*	29820:	gap of unknown length
*	29821	contig of 1489 bp in length
*	29921	contig of 1489 bp in length
*	31409:	gap of unknown length
*	31410	contig of 1798 bp in length
*	31510	contig of 1798 bp in length
*	33307:	gap of unknown length
*	33407:	contig of 2019 bp in length
*	33408	contig of 2019 bp in length
*	35425:	gap of unknown length
*	35427	contig of 1354 bp in length
*	35526:	gap of unknown length
*	35527	contig of 1549 bp in length
*	36881	contig of 1549 bp in length
*	36980:	gap of unknown length
*	38529:	contig of 1517 bp in length
*	38530	contig of 1517 bp in length
*	38629:	gap of unknown length
*	38630	contig of 2271 bp in length
*	40147	contig of 2271 bp in length
*	40247	gap of unknown length
*	40248:	contig of 1675 bp in length
*	42518	contig of 1675 bp in length
*	44292:	gap of unknown length
*	44293:	contig of 2316 bp in length
*	44392:	gap of unknown length
*	44393	contig of 2434 bp in length
*	46709	gap of unknown length
*	46809	contig of 1916 bp in length
*	46809	contig of 1916 bp in length
*	49343	gap of unknown length
*	49343	contig of 2002 bp in length
*	51359	gap of unknown length
*	51359	contig of 1688 bp in length
*	53361	gap of unknown length
*	53461	contig of 1688 bp in length
*	55149	gap of unknown length
*	55149	contig of 1688 bp in length

```

* 55249 58677: contig of 3429 bp in length
* 58678 58777: gap of unknown length
* 58778 60637: contig of 1860 bp in length
* 60638 60737: gap of unknown length
* 60738 63930: contig of 3193 bp in length
* 63931 64030: gap of unknown length
* 64031 66398: contig of 2368 bp in length
* 66399 66498: gap of unknown length
* 66499 69978: contig of 3480 bp in length
* 69979 70078: gap of unknown length
* 70079 72552: contig of 2474 bp in length
* 72553 74753: gap of unknown length
* 74753 77949: contig of 2100 bp in length
* 77949 78050: gap of unknown length
* 78050 80099: contig of 1950 bp in length
* 80099 82883: gap of unknown length
* 82883 82984: contig of 2784 bp in length
* 82984 86241: gap of unknown length
* 86241 90145: contig of 3257 bp in length
* 90145 90245: gap of unknown length
* 90245 92780: contig of 3805 bp in length
* 92780 92881: gap of unknown length
* 92881 97022: contig of 2535 bp in length
* 97022 97123: gap of unknown length
* 97123 99914: contig of 4142 bp in length
* 99914 100014: gap of unknown length
* 100014 103193: contig of 2792 bp in length
* 103193 103292: gap of unknown length
* 103292 107022: contig of 3178 bp in length
* 107022 107122: gap of unknown length
* 107122 110597: contig of 3730 bp in length
* 110597 110697: gap of unknown length
* 110697 116078: contig of 3475 bp in length
* 116078 120578: gap of unknown length
* 120578 120678: contig of 5380 bp in length
* 120678 125200: gap of unknown length
* 125200 125200: contig of 4401 bp in length
* 125200 125200: gap of unknown length
* 125200 125200: contig of 4522 bp in length

```

```

Query Match 92.7% Score 20.4; DB 2; Length 152147;
Best Local Similarity 95.5%; Pred. No. 63;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 AAGAAAATCTAGACAGCAA 22
Db 141943 AAGAAAATGTAGACAGCAA 141964

```

```

RESULT 3
AC117913/c 172497 bp DNA linear HTG 18-JUL-2002
LOCUS Rattus norvegicus clone CH230-35D24, *** SEQUENCING IN PROGRESS
DEFINITION *** 59, unordered pieces.
ACCESSION AC117913
VERSION AC117913.4 GI:21747159
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 172497)
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Bunay,C., Butch,P., Burkelt,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

```

```

COMMENT

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL

```

```

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgeson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlssohn,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louised,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Prims,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,O.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 172497)
Worley,K.C.
Direct Submission
Submitted (11-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 172497)
Worley,K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20258282.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYPM
Center clone name: CH230-35D24
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 118901 bases at least Q40
Consensus quality: 125278 bases at least Q30
Consensus quality: 129335 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* 1 1133: contig of 1133 bp in length
* 1134 1233: gap of unknown length
* 1234 2457: contig of 1224 bp in length
* 2458 2557: gap of unknown length
* 2558 3893: contig of 1336 bp in length
* 3894 3993: gap of unknown length
* 3994 5079: contig of 1086 bp in length
* 5080 5179: gap of unknown length
* 5180 6739: contig of 1560 bp in length
* 6740 8037: gap of unknown length
* 8038 8137: gap of unknown length
* 8138 9770: contig of 1632 bp in length
* 9770 9870: gap of unknown length
* 9870 11174: contig of 1304 bp in length
* 11174 11274: gap of unknown length
* 11274 12328: contig of 1056 bp in length
* 12330 12429: gap of unknown length
* 12430 14363: contig of 1934 bp in length
* 14364 14463: gap of unknown length
* 14464 15857: contig of 1394 bp in length
* 15858 15957: gap of unknown length
* 15958 17164: contig of 1207 bp in length
* 17165 17264: gap of unknown length
* 17265 18407: contig of 1143 bp in length
* 18408 20371: contig of 1864 bp in length
* 20372 20472: gap of unknown length
* 20472 21793: contig of 1321 bp in length
* 21793 21892: gap of unknown length
* 21893 23539: contig of 1647 bp in length
* 23540 23639: gap of unknown length
* 23640 24789: contig of 1150 bp in length
* 24790 24889: gap of unknown length
* 24890 26903: contig of 2014 bp in length
* 26904 27003: gap of unknown length
* 27004 28488: contig of 1485 bp in length
* 28489 28588: gap of unknown length
* 28589 29829: contig of 1241 bp in length
* 29830 29929: gap of unknown length
* 29930 31463: contig of 1533 bp in length
* 31463 31562: gap of unknown length
* 31563 33655: contig of 2093 bp in length
* 33656 33755: gap of unknown length
* 33756 35134: contig of 1379 bp in length
* 35135 35234: gap of unknown length
* 35235 37277: contig of 2043 bp in length
* 37278 37377: gap of unknown length
* 37378 38782: contig of 1405 bp in length
* 38783 38882: gap of unknown length
* 38883 40412: contig of 1530 bp in length
* 40413 40512: gap of unknown length
* 40513 42065: contig of 1553 bp in length
* 42066 42165: gap of unknown length
* 42166 44423: contig of 2258 bp in length
* 44424 44524: gap of unknown length
* 44524 45924: contig of 1401 bp in length
* 45925 46024: gap of unknown length
* 46025 47783: contig of 1755 bp in length
* 47784 47883: gap of unknown length
* 47884 50262: contig of 2379 bp in length
* 50263 50362: gap of unknown length
* 50363 52442: contig of 2080 bp in length
* 52443 52542: gap of unknown length
* 52543 55030: contig of 2488 bp in length
* 55031 55130: gap of unknown length
* 55131 57583: contig of 2453 bp in length
* 57584 57683: gap of unknown length
* 57684 59735: contig of 2052 bp in length
* 59736 59835: gap of unknown length
* 59836 62023: contig of 2188 bp in length
* 62024 62123: gap of unknown length
* 62124 65327: contig of 3204 bp in length

```

```

Query Match          92.7%: Score 20.4; DB 2: Length 172497;
Best Local Similarity 95.5%: Pred. No. 62;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 AAGAAAAATCTGACAGCA 22
Db 45058 AAGAAAAATGTAGACACAA 45037

```

```

RESULT 4
LOCUS AC113320
DEFINITION Mus musculus clone RP23-445E16, WORKING DRAFT SEQUENCE, 11 ordered
pieces.
AC113320
AC113320.2 GI:22474991
VERSION HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS house mouse.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 188795)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-445E16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 188795)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Bouknight, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Lacombe, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Milhova, T.,

```


AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaralunga, H.C., Are, J.R., Banks, T., Barbata, J., Benton, J., Blum, K., Blankenburg, K., Bonin, D., Bouck, J., Bowls, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burich, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carion, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, N., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hayes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homel, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteage, O., Lieu, C., Liu, J., Liu, W., Louissege, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogub, M., Okwuon, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Pirmus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sontheimer, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalón, D., Vinson, R., Wall, L., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, Y., Wu, Y.F., Zhou, Y., Zorrilla, S., Naylor, S.L. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 179767)
Moriarty, K.C.
Direct Submission
Submitted (22-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 179767)
Moriarty, K.C.
Direct Submission
Submitted (30-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 30, 2001 this sequence version replaced gi:13430920.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons

flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Contig length: 179767
Phrap values in estimate: 179551
Average error rate (BCM-Phrap estimate): 4.4907e-06
Fraction of Phrap values less than 40 : 0.00499023
Number of consensus changing edits: 4
Number of N's in consensus : 0
```

```
----- Consensus changing edits -----
Position Original+Context Edited+Context
3 acccaaaa(n)tcgtgacttc atcgcgga(c)tcgacttc
107934 agcctcttt(n)acagagact agcctcttt(c)tcagagact
107935 gctctcttn(n)acagagact gctctcttt(c)tcagagact
159144 atgccaatca(n)cacacatggt atgccaatca(c)cacacatggt
```

----- Distribution of Quality < 40 Bases -----

```
10001
9001
8001
7001
6001
5001
4001
3001
2001
1001
01
5 10 15 20 25 30 35 40
Phrap value Range
```

FEATURES Version: 1.01 gxf.

```
----- Location/Qualifiers -----
source 1..179767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-398021"
55..358
/rpt_family="AluY"
repeat_region complement(635..774)
/rpt_family="MIR"
1202..1410
/rpt_family="AluYb"
repeat_region complement(4054..4428)
/rpt_family="L2"
repeat_region complement(4951..5263)
```

```

repeat_region      /rpt_family="Alusq"
                    /302..7547
repeat_region      /rpt_family="L1PA13"
                    /complement(10047..10274)
repeat_region      /rpt_family="MIR"
                    /complement(10755..10914)
repeat_region      /rpt_family="HAL1"
                    /complement(10947..11385)
repeat_region      /rpt_family="HAL1"
                    /complement(12041..12199)
repeat_region      /rpt_family="AlusB"
                    /1244..12383
repeat_region      /rpt_family="MIR"
                    /complement(13142..13309)
repeat_region      /rpt_family="MER5B"
                    /14874..15137
repeat_region      /rpt_family="AluIo"
                    /complement(16376..16594)
repeat_region      /rpt_family="I2"
                    /17830..19096
misc_feature        /note="Region similar to Homo sapiens mRNA; cDNA
                    DKFZ586L2024 (from clone DKFZ586L2024); partial cds
                    /cds=0..1122 /gb=AL117664 /gi=5912260 /ug=Hs.58419
                    /len=2298"
STS
                    /standard_name="G21620"
                    /db_xref="dbSTS:35458"
                    /complement(19478..20298)
                    /note="Region similar to Homo sapiens rRk-fused gene
                    (TPG), mRNA /cds=(18,1220) /gb=NM_006070 /gi=5174718
                    /ug=Hs.250897 /len=1677"
                    /standard_name="N1B1880"
                    /22027..22107
repeat_region      /rpt_family="MIR"
                    /22097..22217
repeat_region      /rpt_family="MER5B"
                    /complement(22479..22694)
repeat_region      /rpt_family="MER58A"
                    /23342..23375
repeat_region      /rpt_family="An_rich"
                    /complement(25239..25373)
repeat_region      /rpt_family="MIR"

Query Match      88.2% Score 19.4; DB 9; Length 179767;
Best Local Similarity 95.2% Pred No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAGAAAAATCTAGACAGCA 21
Db 4959 AAGAAAAATCTAGACAGCA 4939

RESULT 6
AC068763 180638 bp DNA linear PRI 25-JUL-2002
LOCUS Homo sapiens 3 BAC RP11-133K20 (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
ACCESSION AC068763
VERSION AC068763.11 GI:14669935
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 180638)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-osman,F.R., Allen,C.,
Alshrocks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
JOURNAL

```

COyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
 He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M.,
 Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Huylk,S.,
 Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvay,J.,
 Kovar,C., Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
 Louised,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Kawhney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
 Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
 Nickerson,E., Nwokenkwo,S., Ogub,M., Okunolu,G., Orangunye,N.,
 Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
 Pickens,R., Prins,E., Pu,L.L., Quiles,M., Ren,Y., Rivers,M.,
 Scott,G., Shen,H., Shooshairi,N., Sisson,I., Sodergren,E.,
 Rojas,A., Rojudo-Kan,I., Rolfe,M., Ruiz,S., Severy,G., Scherer,S.,
 Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Sytek,A.,
 Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
 Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
 Veyr,V., Villalob,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S.,
 Warren,R., Washington,S., Wallington,C., Williams,G.,
 Williams,A., Wlezyk,R., Woodson,S., Worley,K., Wu,C., Wu,Y.,
 Wu,Y.F., Zhou,J., Zorilla,S., Zorilla,S.L., and Gibbs,R.

Unpublished
 2 (bases 1 to 180638)
 Morley,K.C.
 Direct Submission
 Submitted (09-MAY-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 180638)
 Morley,K.C.
 Direct Submission
 Submitted (11-JUL-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 180638)
 Morley,K.C.
 Direct Submission
 Submitted (02-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 180638)
 Morley,K.C.
 Direct Submission
 Submitted (25-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2001 this sequence version replaced gi:13493008.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 features listing.

ANNOTATION OF FEATURES:
 STSs are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Contig length: 180638
Phrap values in estimate: 179362
Average error rate (BCM-Phrap estimate): 1.00944e-05
Fraction of Phrap values less than 40 : 0.00994637
Number of consensus changing edits: 49
0
```

```
----- Consensus changing edits -----
Position Original+Context Edited+Context
11715 tgttataga(n)aaagagctc ttttataga(a)aaagagctc
11767 tatgataga(n)acgtgcgtc tatgataga(a)acgtgcgtc
13821 atctagaaga(n)atgataaat atctagaaga(a)atgataaat
13822 tctagaaga(n)tgataaat tctagaaga(a)tgataaat
29835 gattgctta(n)gattgctta gattgctta(a)gattgctta
49503 acttcagta(n)acttcagta acttcagta(a)acttcagta
51340 tccatcatt(n)gattcctac tccatcatt(a)gattcctac
52598 ggaataaat(n)catgtacaa ggaataaat(a)catgtacaa
60642 acttaaaat(n)aaatttcgg acttaaaat(a)aaatttcgg
60920 caacaagag(c)taactcatt caacaagag(c)taactcatt
60921 aacaagagct(t)taactcatt aacaagagct(a)taactcatt
60922 agatgcgct(a)actcattc agatgcgct(a)actcattc
64475 atctagctc(n)ctgcactca atctagctc(a)ctgcactca
73510 tcaagagat(n)tcctgcctc tcaagagat(a)tcctgcctc
73543 tctgctc(n)ctcctcaaat tctgctc(a)ctcctcaaat
73566 tcccaaat(n)ctggattac tcccaaat(a)ctggattac
73567 tcccaaat(n)tgattata tcccaaat(a)tgattata
73704 ccacatgc(n)gctaatatt ccacatgc(a)gctaatatt
73721 gttggagat(n)acagcgctg gttggagat(a)acagcgctg
73722 gttggagat(n)acagcgctc gttggagat(a)acagcgctc
73748 tggacacac(n)tgctgcctc tggacacac(a)tgctgcctc
73750 tttctatta(n)tttgatagt tttctatta(a)tttgatagt
73756 accagctcct(n)gattatga accagctcct(a)gattatga
101317 ctgtagctt(n)taaacctca ctgtagctt(a)taaacctca
109869 tttttttt(n)gagagagctc tttttttt(a)gagagagctc
109869 gttggagag(n)gagtgagag gttggagag(a)gagtgagag
110872 gttggagag(n)gagtgagag gttggagag(a)gagtgagag
113503 gttgcctc(n)ncattttaa gttgcctc(a)ncattttaa
125301 ttgcctc(n)caattttaa ttgcctc(a)caattttaa
128519 aaatacct(n)ataattatc aaatacct(a)ataattatc
13447 ctgataatc(n)gcttttttc ctgataatc(a)gcttttttc
13447 ttcggagat(n)tcctcctaa ttcggagat(a)tcctcctaa
13442 cagctgctt(n)ctagcctaa cagctgctt(a)ctagcctaa
137769 cctgctag(n)gtggtgctc cctgctag(a)gtggtgctc
```

```
137774 ctgagntg(n)gctcagcct ctgagntg(a)gctcagcct
111481 tagacaatg(n)actttaagt tagacaatg(a)actttaagt
151926 atactcaga(n)aaatccaaa atactcaga(a)aaatccaaa
180507 taccagag(n)gmmmmmtt taccagag(a)gmmmmmtt
180509 accaggtg(n)nnmmmtat accaggtg(a)nnmmmtat
180510 ccaagttg(n)nnmmmtat ccaagttg(a)nnmmmtat
180511 caggtgag(n)nnmmmtat caggtgag(a)nnmmmtat
180512 aggtgag(n)nnmmmtat aggtgag(a)nnmmmtat
180513 ggtgagag(n)nnmmmtat ggtgagag(a)nnmmmtat
180514 gttgagag(n)nnmmmtat gttgagag(a)nnmmmtat
180515 tggagagag(n)nnmmmtat tggagagag(a)nnmmmtat
180516 gggagagag(n)nnmmmtat gggagagag(a)nnmmmtat
180559 tttatc(n)aaatattt tttatc(a)aaatattt
```

----- Distribution of Quality < 40 Bases -----

```
1000|
900|
800|
700|
600|
500|
400|
300|
200|
100|
0|
* * * * *
5 10 15 20 25 30 35 40
Phrap Value Range
```

FEATURES Location/Qualifiers

Query Match 88.2%; Score 19.4; DB 9; Length 180638;
Best Local Similarity 95.2%; Pred. No. 1.6e-02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AACGAAAAATCTGACAGCA 21
Db 163480 AACGAAAAATCTGACAGCA 163500

RESULT 7
AC027080 185841 bp DNA linear HTG 19-APR-2000

LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-548023 map 3, WORKING DRAFT
AC027080.2 GI:7596901
VERSION
AC027080.2 GI:7596901
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.

REFERENCE
1 (bases 1 to 185841)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Biren, B., Linton, L., Nussbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 185841)

REFERENCE
1 (bases 1 to 185841)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Biren, B., Linton, L., Nussbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 185841)

REFERENCE
1 (bases 1 to 185841)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Biren, B., Linton, L., Nussbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 185841)

TITLE
JOURNAL
COMMENT

Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margolis, N.,
McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McNeeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 19, 2000 this sequence version replaced g1:7329443.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence-submissions@genome.wi.mit.edu

Project Information

Center project name: L8201

Center clone name: 548_O_23

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 172210 bases at least Q40

Consensus quality: 178990 bases at least Q30

Consensus quality: 181814 bases at least Q20

Insert size: 189000; agarose-ef

Insert size: 183741; sum-of-contents

Quality coverage: 4.0 in Q20 bases; agarose-ef

Quality coverage: 4.1 in Q20 bases; sum-of-contents

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1458: contig of 1458 bp in length
* 1459 1558: gap of 100 bp
* 1559 3252: contig of 1694 bp in length
* 3253 3352: gap of 100 bp
* 3353 5248: contig of 1896 bp in length
* 5249 5348: gap of 100 bp
* 5349 8165: contig of 2817 bp in length
* 8166 8265: gap of 100 bp
* 8266 11611: contig of 3346 bp in length
* 11612 11711: gap of 100 bp
* 11712 16128: contig of 4417 bp in length
* 16129 16228: gap of 100 bp
* 16229 22113: contig of 5885 bp in length
* 22114 22213: gap of 100 bp
* 22214 28574: contig of 6361 bp in length
* 28575 28674: gap of 100 bp
* 28675 34052: contig of 5378 bp in length
* 34053 34152: gap of 100 bp
* 34153 41388: contig of 7236 bp in length
* 41389 41488: gap of 100 bp
* 41489 50939: contig of 9451 bp in length
* 50940 51039: gap of 100 bp
* 51040 61135: contig of 10096 bp in length
* 61136 61235: gap of 100 bp
* 61236 69782: contig of 8547 bp in length
* 69783 69882: gap of 100 bp
* 69883 77230: contig of 7348 bp in length

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

77231 77330: gap of 100 bp
* 77331 86536: contig of 9206 bp in length
* 86537 86636: gap of 100 bp
* 86637 98162: contig of 11526 bp in length
* 98163 98262: gap of 100 bp
* 98263 109460: contig of 11198 bp in length
* 109461 109560: gap of 100 bp
* 109561 121285: contig of 11725 bp in length
* 121286 121385: gap of 100 bp
* 121386 135542: contig of 14157 bp in length
* 135543 135642: gap of 100 bp
* 135643 147139: contig of 11497 bp in length
* 147140 147239: gap of 100 bp
* 147240 161924: contig of 14685 bp in length
* 161925 162024: gap of 100 bp
* 162025 185841: contig of 2817 bp in length.
Location/Qualifiers
1. 185841
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-548023"
/clone_lib="RP11 Human Male BAC"
1. 1458
/note="assembly-fragment"
/note="assembly-fragment"
1559..3252
/note="assembly-fragment"
3353..5248
/note="assembly-fragment"
5349..8165
/note="assembly-fragment"
clone_end:77
vector_side:right"
8266..11611
/note="assembly-fragment"
11712..16128
/note="assembly-fragment"
16229..22113
/note="assembly-fragment"
22214..28574
/note="assembly-fragment"
28675..34052
/note="assembly-fragment"
34153..41388
/note="assembly-fragment"
41489..50939
/note="assembly-fragment"
51040..61135
/note="assembly-fragment"
61236..69782
/note="assembly-fragment"
69883..77230
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
77331..86536
/note="assembly-fragment"
86637..98162
/note="assembly-fragment"
98263..109460
/note="assembly-fragment"
109561..121285
/note="assembly-fragment"
121386..135542
/note="assembly-fragment"
135643..147139
/note="assembly-fragment"
147240..161924
/note="assembly-fragment"
162025..185841
/note="assembly-fragment"

BASE COUNT 57091 a 34837 c 33673 g 58130 t 2110 others

ORIGIN

Query Match 88.2% Score 19.4; DB 2; Length 185841;
 Best Local Similarity 95.2% Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACAGCA 21
 |||||
 Db 127736 AAGAAAAATCTAGACAGCA 127756

RESULT 8
 AC130437/c 201416 bp DNA 11near HTG 10-AUG-2002
 LOCUS Homo sapiens chromosome 3 clone RP11-740L19, *** SEQUENCING IN
 DEFINITION PROGRESS *** 10 unordered pieces.
 AC130437
 AC130437.1 GI:22203194
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 201416)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbara,J., Benton,J., Blumage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorielli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homel,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kretovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lousseged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Mettzer,M.,
 Miner,G., Miner,Z., Newton,J., Newton,N., Nguyen,A., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokkwo,S., Ogulu,M., Okunodu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Qules,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shoshitani,N., Sisson,I.,
 Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Swatek,A., Tabor,P., Tameis,A., Tameis,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Glbbs,R.

TITLE Direct Submission
 JOURNAL Unpublished
 AUTHORS 2 (bases 1 to 201416)
 TITLE Direct Submission
 JOURNAL Submitted (10-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 COMMENT ----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: HDIU
 Center clone name: RP11-740L19

Summary Statistics
 Chemistry: Dye-Primer Bodipies
 Chemistry: Dye-terminator Big Dye; Infinity% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 200930 bases at least Q40
 Consensus quality: 216627 bases at least Q30
 Consensus quality: 233347 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 10 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 2055: contig of 2055 bp in length
 2056 2155: gap of unknown length
 2156 4275: contig of 2120 bp in length
 4276 4375: gap of unknown length
 4376 9623: contig of 5248 bp in length
 9624 9723: gap of unknown length
 9724 26424: contig of 16700 bp in length
 26424 26522: gap of unknown length
 26522 42616: contig of 16093 bp in length
 42616 42716: gap of unknown length
 42716 59134: contig of 16418 bp in length
 59134 59234: gap of unknown length
 59234 77937: contig of 18703 bp in length
 77937 78037: gap of unknown length
 78037 98933: contig of 20856 bp in length
 98933 99033: gap of unknown length
 99033 148069: contig of 49035 bp in length
 148069 148168: gap of unknown length
 148168 201416: contig of 53248 bp in length.

Location/Qualifiers
 1. 201416
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-740L19"

BASE COUNT 63751 a 38726 c 37375 g 60653 t 911 others

ORIGIN

Query Match 88.2% Score 19.4; DB 2; Length 201416;
 Best Local Similarity 95.2% Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACAGCA 21
 |||||
 Db 174985 AAGAAAAATCTAGACAGCA 174965

RESULT 9
 AC022912 218859 bp DNA 11near HTG 26-MAY-2000
 LOCUS Homo sapiens clone RP11-740L19, WORKING DRAFT SEQUENCE, 34
 DEFINITION UNORDERED PIECES.
 AC022912
 AC022912.3 GI:7596818
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia: Eutheria: Primates; Catarrhini: Homidae: Homo.
1 (bases 1 to 218859)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-740L19
Unpublished
2 (bases 1 to 218859)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckely, R., Beda, F., Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G., Castle, A., Choe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArnell, K., Dewar, K., Domino, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardina, S., Grant, G., Hagos, B., Heath, A., Horton, L., Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lepocky, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Menes, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Rhmann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On April 19, 2000 this sequence version replaced gi:1158108.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L6286

Center clone name: 740.L.19

Summary Statistics

Sequencing vector: M13, M7815, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 186472 bases at least Q40
Consensus quality: 198582 bases at least Q30
Consensus quality: 206779 bases at least Q20
Insert size: 185000; agarose-
Insert size: 215559; sum-of-
Quality coverage: 4.5 in Q20 bases; agarose-
Quality coverage: 3.9 in Q20 bases; sum-of-
contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1210: contig of 1210 bp in length
* 1211 1310: gap of 100 bp
* 1311 2361: contig of 1051 bp in length
* 2362 2461: gap of 100 bp
* 2462 3783: contig of 1322 bp in length
* 3784 3883: gap of 100 bp
* 3884 5069: contig of 1186 bp in length
* 5070 5169: gap of 100 bp
* 5170 6265: contig of 1096 bp in length
* 6266 6365: gap of 100 bp
* 6366 7846: contig of 1481 bp in length
* 7847 7946: gap of 100 bp
* 7947 8992: contig of 1046 bp in length
* 8993 9092: gap of 100 bp
* 9093 10112: contig of 1020 bp in length

* 10113 10212: gap of 100 bp
* 10213 11548: contig of 1336 bp in length
* 11549 11648: gap of 100 bp
* 11649 12946: contig of 1298 bp in length
* 12947 13046: gap of 100 bp
* 13047 14927: contig of 1881 bp in length
* 14928 15027: gap of 100 bp
* 15028 16366: contig of 1339 bp in length
* 16367 16466: gap of 100 bp
* 16467 17996: contig of 1530 bp in length
* 17997 18096: gap of 100 bp
* 18097 20861: contig of 2765 bp in length
* 20862 20961: gap of 100 bp
* 20962 22546: contig of 1585 bp in length
* 22547 22646: gap of 100 bp
* 22647 25390: contig of 2744 bp in length
* 25391 25490: gap of 100 bp
* 25491 28906: contig of 3416 bp in length
* 28907 29006: gap of 100 bp
* 29007 33315: contig of 4309 bp in length
* 33316 33415: gap of 100 bp
* 33416 35211: contig of 1796 bp in length
* 35212 35311: gap of 100 bp
* 35312 40323: contig of 5012 bp in length
* 40324 40423: gap of 100 bp
* 40424 44661: contig of 4238 bp in length
* 44662 44761: gap of 100 bp
* 44762 50183: contig of 5422 bp in length
* 50184 50283: gap of 100 bp
* 50284 56649: contig of 6266 bp in length
* 56650 63660: contig of 6911 bp in length
* 63661 63660: gap of 100 bp
* 63661 73767: contig of 10107 bp in length
* 73768 73867: gap of 100 bp
* 73868 84412: contig of 10545 bp in length
* 84413 84512: gap of 100 bp
* 84513 95985: contig of 11473 bp in length
* 95986 96085: gap of 100 bp
* 96086 109032: contig of 12947 bp in length
* 109033 109132: gap of 100 bp
* 109133 120785: contig of 11653 bp in length
* 120786 120885: gap of 100 bp
* 120886 132423: contig of 11538 bp in length
* 132424 132523: gap of 100 bp
* 132524 151790: contig of 19267 bp in length
* 151791 151890: gap of 100 bp
* 151891 170984: contig of 19094 bp in length
* 170985 171084: gap of 100 bp
* 171085 193704: contig of 22620 bp in length
* 193705 218859: gap of 100 bp
* 218859 218859: contig of 25055 bp in length.

FEATURES

source

misc_feature
1..1210
/note="assembly-fragment"
misc_feature
1311..2361
/note="assembly-fragment"
misc_feature
2462..3783
/note="assembly-fragment"
misc_feature
3884..5069
/note="assembly-fragment"
misc_feature
5170..6265
/note="assembly-fragment"
misc_feature
6366..7846
/note="assembly-fragment"
misc_feature
7947..8992
/note="assembly-fragment"
misc_feature
9093..10112
/note="assembly-fragment"

```

misc_feature      /note="assembly_fragment"
10213..11548
/note="assembly_fragment"
misc_feature      11649..12946
/note="assembly_fragment"
misc_feature      13047..14927
/note="assembly_fragment"
misc_feature      15028..16366
/note="assembly_fragment"
misc_feature      16467..17996
/note="assembly_fragment"
misc_feature      18097..20861
/note="assembly_fragment"
misc_feature      20962..22546
/note="assembly_fragment"
misc_feature      22647..25390
/note="assembly_fragment"
misc_feature      25491..28906
/note="assembly_fragment"
misc_feature      29007..33315
/note="assembly_fragment"
misc_feature      33416..35211
/note="assembly_fragment"
misc_feature      clone_end:77
vector_side:right"
misc_feature      35312..40323
/note="assembly_fragment"
misc_feature      40424..44661
/note="assembly_fragment"
misc_feature      44762..50183
/note="assembly_fragment"
misc_feature      50284..56549
/note="assembly_fragment"
misc_feature      56650..63560
/note="assembly_fragment"
misc_feature      63661..73767
/note="assembly_fragment"
misc_feature      73868..84412
/note="assembly_fragment"
misc_feature      84513..95985
/note="assembly_fragment"
misc_feature      96086..109032
/note="assembly_fragment"
misc_feature      109133..120785
/note="assembly_fragment"

Query Match      88.2%: Score 19.4: DB 2: Length 218859;
Best Local Similarity 95.2%: Pred NO.1.6e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAGCAAAAATCTAGACAGCA 21
DB 146022 AAGCAAAAATCTAGACAGCA 146042

RESULT 10
AC093623/c      274349 bp      DNA      linear      HTG 07-SEP-2001
LOCUS      Homo sapiens chromosome UNK clone CTD-2335D13, *** SEQUENCING IN
DEFINITION
AC093623
AC093623.1      GI:15487445
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 274349)
AUTHORS
Waterston,R.H.
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 274349)
AUTHORS
Waterston,R.H.

```

TITLE JOURNAL COMMENT

```

Direct Submission
Submitted (07-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Project name: H_M82335D13
Center project name: H_M82335D13
----- Summary Statistics -----
Sequencing vector: M13; 1%
Sequencing vector: plasmid; 99%
Chemistry: Dye-terminator Big Dye; 99% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 248392 bases at least Q40
Consensus quality: 257485 bases at least Q30
Consensus quality: 262193 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1390 1389: contig of 1389 bp in length
1490 1489: gap of unknown length
2557 2556: contig of 1067 bp in length
2657 2656: gap of unknown length
3865 3865: contig of 1209 bp in length
3966 3965: gap of unknown length
5038 5038: contig of 1073 bp in length
5139 5138: gap of unknown length
6763 6762: contig of 1624 bp in length
6863 6862: gap of unknown length
8029 8029: contig of 1167 bp in length
8129 8129: gap of unknown length
8130 8129: contig of 1116 bp in length
9245 9245: gap of unknown length
9346 9345: gap of unknown length
10442 10441: contig of 1096 bp in length
10542 10541: gap of unknown length
11789 11788: contig of 1247 bp in length
11889 11888: gap of unknown length
13219 13219: contig of 1331 bp in length
13320 13319: gap of unknown length
14328 14328: contig of 1009 bp in length
14429 14428: gap of unknown length
15578 15578: contig of 1150 bp in length
15678 15678: gap of unknown length
15679 15678: contig of 1564 bp in length
17243 17242: gap of unknown length
17343 17342: gap of unknown length
19057 19056: contig of 1714 bp in length
19157 19156: gap of unknown length
20356 20355: contig of 1199 bp in length
20455 20455: gap of unknown length
21888 21888: contig of 1433 bp in length
21989 21988: gap of unknown length
23748 23748: contig of 1760 bp in length
23849 23848: gap of unknown length
25791 25791: contig of 1943 bp in length
25891 25891: gap of unknown length
29014 29014: contig of 3123 bp in length
29115 29114: gap of unknown length
30653 30653: contig of 1539 bp in length
30754 30753: gap of unknown length
32579 32579: contig of 1826 bp in length
32680 32679: gap of unknown length
34693 34693: contig of 2014 bp in length
34793 34793: gap of unknown length

```

34744	36794:	contig of 2001 bp in length
36795	36694:	gap of unknown length
36695	39286:	contig of 2392 bp in length
39287	39386:	gap of unknown length
42071	42070:	contig of 2684 bp in length
42171	42170:	gap of unknown length
45224	45523:	contig of 3353 bp in length
45624	45623:	gap of unknown length
47929	47928:	contig of 2305 bp in length
48029	48028:	gap of unknown length
50379	50378:	contig of 2350 bp in length
50479	50478:	gap of unknown length
52146	52145:	contig of 1667 bp in length
52246	52245:	gap of unknown length
54934	54934:	contig of 2669 bp in length
55034	55034:	gap of unknown length
55035	55039:	contig of 2505 bp in length
57540	57539:	gap of unknown length
57640	60553:	contig of 2394 bp in length
60334	60633:	gap of unknown length
60634	63866:	contig of 3223 bp in length
63867	63966:	gap of unknown length
63967	67394:	contig of 3428 bp in length
67395	67494:	gap of unknown length
71251	71251:	contig of 3757 bp in length
71252	71351:	gap of unknown length
71352	75513:	contig of 4162 bp in length
75513	75613:	gap of unknown length
75614	79500:	contig of 3887 bp in length
79501	79600:	gap of unknown length
79601	85965:	contig of 6355 bp in length
85965	86055:	gap of unknown length
86056	90886:	contig of 4841 bp in length
90887	90996:	gap of unknown length
90997	95402:	contig of 4406 bp in length
95403	95502:	gap of unknown length
95503	100307:	contig of 4805 bp in length
100308	100407:	gap of unknown length
100408	105113:	contig of 4706 bp in length
105114	105216:	gap of unknown length
105216	109576:	contig of 4363 bp in length
109577	109676:	gap of unknown length
114723	114722:	contig of 5046 bp in length
114723	114822:	gap of unknown length
114823	119838:	contig of 5016 bp in length
119839	119938:	gap of unknown length
119939	128278:	contig of 8340 bp in length
128279	128378:	gap of unknown length
128379	137203:	contig of 8825 bp in length
137204	137303:	gap of unknown length
137304	170951:	contig of 33668 bp in length
170952	171051:	gap of unknown length
171052	274349:	contig of 103298 bp in length

```
misc_feature      9346. 10441 /note="assembly_name:Config32"
misc_feature      10542. 11788 /note="assembly_name:Config34"
misc_feature      11889. 13219 /note="assembly_name:Config35"
misc_feature      13320. 14328 /note="assembly_name:Config36"
misc_feature      14429. 15578 /note="assembly_name:Config37"
misc_feature      15679. 17242 /note="assembly_name:Config39"
misc_feature      17343. 19056 /note="assembly_name:Config41"
misc_feature      19157. 20355 /note="assembly_name:Config42"
misc_feature      20456. 21888 /note="assembly_name:Config43"
misc_feature      21989. 23748 /note="assembly_name:Config44"
misc_feature      23849. 25791 /note="assembly_name:Config45"
misc_feature      25892. 29014 /note="assembly_name:Config46"
misc_feature      29115. 30653 /note="assembly_name:Config47"
misc_feature      30754. 32579 /note="assembly_name:Config49"
misc_feature      32680. 34693 /note="assembly_name:Config50"
misc_feature      34794. 36794 /note="assembly_name:Config51"
misc_feature      36895. 39286 /note="assembly_name:Config52"
misc_feature      39387. 42070 /note="assembly_name:Config53"
misc_feature      42171. 45523 /note="assembly_name:Config54"
```

Query Match	88.28;	Score 19.4;	DB 2;	Length 274349;
Best Local Similarity	95.28;	Pred. No. 1.5e+02;		
Matches	20;	Conservative	0;	Mismatches 1;
			Indels	0;
			Gaps	0;

[illegible]

COMMENT

(E-mail: tsesaki@nlas.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contig. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source

1. 113940
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:3947"
/chromosome="9"
/clone="OJ1506_A04"

BASE COUNT 33810 a 24133 c 23762 g 32028 t 207 others
ORIGIN

Query Match 86.4% Score 19; DB 2; Length 113940;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACGANG 19

Db 112166 AAGAAAAATCTAGACGANG 112148

RESULT 12 AF012746 2052 bp mRNA linear VRT 24-OCT-2000
LOCUS AF012746.1 GI:2331258
DEFINITION Danto olfactory receptor protein 13.1 mRNA, complete cds.
ACCESSION AF012746.1
VERSION AF012746.1
KEYWORDS Danto olfactory receptor protein 13.1 mRNA, complete cds.
SOURCE Danto olfactory receptor protein 13.1 mRNA, complete cds.
ORGANISM Danto olfactory receptor protein 13.1 mRNA, complete cds.

REFERENCE 1 (bases 1 to 2052)
AUTHORS Barth, A.L., Justice, N.J., and Ngai, J.
TITLE Asynchronous onset of odorant receptor expression in the developing zebrafish olfactory system

JOURNAL Neuron 16 (1), 23-34 (1996)

MEDLINE 96158919

PUBMED 8562087

REFERENCE 2 (bases 1 to 2052)

AUTHORS Barth, A.L., Dugas, J.C., and Ngai, J.

TITLE Noncoordinate expression of odorant receptor genes tightly linked in the zebrafish genome

JOURNAL Neuron 19 (2), 359-369 (1997)

MEDLINE 97436752

PUBMED 9292725

REFERENCE 3 (bases 1 to 2052)

AUTHORS Barth, A.L.

TITLE Direct Substitution

JOURNAL Submitted (07-DEC-1995) Molecular and Cellular Biology, University of California, Berkeley, Km. 265 USA, Berkeley, CA 94720, USA

REFERENCE 4 (bases 1 to 2052)

AUTHORS Barth, A.L., Dugas, J.C., and Ngai, J.

TITLE Direct Substitution

JOURNAL Submitted (08-JUL-1997) Molecular and Cellular Biology, University of California, Berkeley, Km. 265 USA, Berkeley, CA 94720, USA

REMARK Sequence updated by submitter
On Oct 24, 2000 this sequence version replaced g1:1151128.

FEATURES

Location/Qualifiers
1..2052

CDS

/organism="Danto olfactory"
/db_xref="taxon:7955"
/tissue_type="olfactory"
17..1003
/note="Previously named olfactory receptor protein 13"

/codon_start=1
/product="olfactory receptor protein 13.1"

/protein_id="AAC60253.1"

/db_xref="GI:2331259"

/translation="MSAGNISVEFFIVGPGIOLGCPDSVAALMFVPCVLAAG
TELFISERKSLKPMYVILNLVASPILESTTLRITRIKRLFGDGGYSFCCFTOM
FEVHFGVNAFLIAAFDRYVAICNPLRPVIVIKWNIYICLSWIAVNTPLM
VIRVPLPYCHENTITOCYCDHISITSLACTNAPYSIPAFVLAVALPLAFVFS
VCATILAVLRISSTQARKTFTCSPOLIITIALYFLPFCITVLSNIGVFSIDLIA
IMMYSLFPPMFNFIYCLRTKDYKCEVLRKLSINVKRNKRVDSVSS"

BASE COUNT 638 a 319 c 293 g 802 t
ORIGIN

Query Match 85.5% Score 18.8; DB 5; Length 2052;
Best Local Similarity 90.9%; Pred. No. 6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACGACAA 22

Db 1472 AAGAAAAATCTAGACGACAA 1451

RESULT 13 SCYPL252C 2209 bp DNA linear PLN 07-AUG-1997
LOCUS SCYPL252C
DEFINITION S. cerevisiae chromosome XVI reading frame ORF YPL252c.
ACCESSION Z73608 U00094
VERSION Z73608.1 GI:1370516
KEYWORDS Saccharomyces cerevisiae.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae.

REFERENCE 1 (bases 1 to 2209)
AUTHORS Pohl, T.M.
TITLE Unpublished

JOURNAL 2 (bases 1 to 2209)

AUTHORS MIPS.

TITLE Direct Substitution

JOURNAL Submitted (28-MAY-1996) Data collected by MIPS on behalf of the

European yeast chromosome XVI sequencing project. MIPS at the

Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152

Martinsried, FRG; E-mail: Mewes@mps.embneth.org

Location/Qualifiers
1..2209

/organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

/chromosome="XVI"

/complement(1..907)

/note="ORF YPL252c"

/codon_start=1

/protein_id="CAA9797.1"

/db_xref="GI:2326844"

/translation="MASOONKHAFLSKRIRFHPNDVSSSKSRNLMDITNTNMGNS
RSSSMKSLALPPVKDSEPPSVRSASLINSKIKDKRDKRTRRORHLRQLTIC
EREIKTFKRDNLKRSFELKYSKQAVYIKOVDLTONLSKSGERADLRKNSAL
ATLOAEIDONLTKROSEQLYNNKLTIWENEDLIMENVEHTEISQLKRTLOE
UNIMNANOKOULEQVNHESQKRDFAFRKAEKLSMENLTKNHRRLDQIATLQSE
SEKLRKELMDIDROAEYSEONISEIN"

/complement(1264..1782)

/note="ORF YPL252c"

/codon_start=1

/protein_id="CAA9797.1"

/db_xref="GI:1370517"

/db_xref="SPTREMBL:Q12184"

/translation="WLKIVTQAGHTARISNIAHLNLTSPSLTRTTTTRFPLPFS
SFLNHLKRPKEELKTFILKDGSKTYVECEGTETIIDIAGHNLDMEGACGGSC

ACSTCHVIVDPDYDALPEPEDDENDMDLAYGLTETSRLGQIKMSKDIDGIRVALP
OMFRVNNNDPS"
1526..1828
/note="orf yPL251w"
/codon_start=1
/protein_id="CAA97976.1"
/db_xref="gi:1370518"
/db_xref="SPTREMBL:Q12058"
/translation="MSRMVSPSQTSTYVWEPSEFRITKVISSPGFGEKMPMKNEV
EKGRNLVYVAVSPDGEVKKRCAMPEILAVCPARVITFISGGLSIDTSAYFV"

BASE COUNT 600 a 362 c 474 g 773 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 8; Length 2209;
Best Local Similarity 90.9%; Pred. No. 6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGAAAAATCTAGACAGCAA 22
|||||
Db 541 AAGAAAAATCCAACAGCAA 520

RESULT 14
SCYPL253C/ 2411 bp DNA linear PLN 11-AUG-1997
LOCUS SCYPL253C
DEFINITION S.cerevisiae chromosome XVI reading frame ORF YPL253C.
ACCESSION Z73609.000094
VERSION Z73609.1 GI:1370519
KEYWORDS
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 2411)
AUTHORS Pohl,T.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2411)
AUTHORS MIPS.
JOURNAL MIPS.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1996) Data collected by MIPS on behalf of the
European yeast chromosome XVI sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.embl.net.org

FEATURES
Location/Qualifiers
1..2411
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="XVI"
complement(112..2055)
/note="orf yPL253c"
/codon_start=1
/protein_id="CAA97978.1"
/db_xref="gi:1370520"
/db_xref="SPTREMBL:Q12045"
/translation="MASQONKHAFLSKNRIFHPNDVSSSKSRNMDITNTTNGMS
RPSMKSIALPVPKDSFSPVSASINIMSKIKDLKRODKIRFQHTLRITOLIEC
EREIKTIKRPDLNKSRELYKKKSKQAYLKQVRODLQONPNSKDEGADILKKNSAL
ATLOAELDONLILKROSOELYNKKILFVNEQLQIMVNEPDHETTELISQIKTLOE
LNIWMANLQONLEROVNHESQLRKDFIAFKKALKSMENLTKNHRLLDQIATLOE
SEKLEKIMIDROAEYSEONISEINENIKOLELANNPLISKSLSQSDLEHLQONQE
NLEKMAKOEKFEYNDYNTVEKELLSRRLSNIIEQGTMRCAVAYMEONLEPNI
DYENGVTGGLSEHVYKFNRYIPLKVSDEKFTOEYSYHDMCLNOKKNPNLSIST
TPHGSLSRESLILFAEKDITIKQYVITLQFVLSDDERSDMLDYSINDKDSIKLK
FEKHSISLDSKLVITENGLEDPLNFCSCDEHPNIPHSQKMIITVOFPPRDSKDGND
PVVDVFIETLNNKSIQEDKSIKFKESCETPIALVLKILSDTKSFLLNLNDSKN
VKKLITSEVOTOLCKRKKLT"

misc-feature
1138..1148
/note="ARS-consensus"

BASE COUNT 654 a 380 c 456 g 921 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 8; Length 2411;
Best Local Similarity 90.9%; Pred. No. 5.9e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGAAAAATCTAGACAGCAA 22
|||||
Db 1689 AAGAAAAATCCAACAGCAA 1668

RESULT 15
SC38RCXVI/ 37808 bp DNA linear PLN 21-OCT-1996
LOCUS SC38RCXVI
DEFINITION S.cerevisiae DNA (chromosome XVI; 38 Kb).
ACCESSION Z67751
VERSION Z67751.1 GI:1061234
KEYWORDS gal4 protein; H(+)-transporting ATPase; HSP82 protein; HSP90
protein; SRP8 protein; SUI3 protein.
SOURCE Saccharomyces cerevisiae.
ORGANISM Saccharomyces cerevisiae.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 37808)
AUTHORS Pohl,T.M.
JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 37808)
AUTHORS Pohl,T.M.
JOURNAL Submitted (09-NOV-1995) Thomas M. Pohl, GATC GmbH,
Fritz-Arnold-Str. 23, Konstanz, 78467, Germany

FEATURES
Location/Qualifiers
1..37808
/organism="Saccharomyces cerevisiae"
/strain="alphas288C"
/db_xref="taxon:4932"
/chromosome="XVI"
/clone_lib="cosmid library (A. Goffeau, Belgium)"
870..2336
/codon_start=1
/product="putative protein"
/protein_id="CAA91590.1"
/db_xref="gi:1061235"
/db_xref="SPTREMBL:Q12060"
/translation="MSAIOAPAPKLOPTYPAPASPASTNAYMKPGLIGSPAVSNHTEP
NNGNNEPAPGCPNORIDLGAMTELETSILGKESWTYKAOIISLFIKLSKELSN
LELVSPSASLEKSNTHHSIVRLNHOILGIFANSIRENPGRNGESSMGESG
SNPNPKLRLKRNKSQLEVRKILVMSIPLDNRNKKITTEACKRGFTICSVQAL
NNIPKIPITVNGDARCVELISLADQYLKNIIEFTIDTVYRRKKSVDYDLNLSG
VGIAREHGLVGDARCVELISLADQYLKNIIEFTIDTVYRRKKSVDYDLNLSG
YKVSSEMAADKRDARIKQDLDKDECDADAKSINNNGSSKDIQISMSITKAG
FAVNELEHNRITISLTNEDIDYSLISFNVLVEPSCSYALTNLGLVNDDELVDMKSN
DLPPLNEKPTPTPDERNCTRHELMWLLKGLTED"

CDS
complement(2446..4391)
/codon_start=1
/product="putative protein"
/protein_id="CAA91591.1"
/db_xref="gi:1061236"
/db_xref="SPTREMBL:Q12045"
/translation="MASQONKHAFLSKNRIFHPNDVSSSKSRNMDITNTTNGMS
RPSMKSIALPVPKDSFSPVSASINIMSKIKDLKRODKIRFQHTLRITOLIEC
EREIKTIKRPDLNKSRELYKKKSKQAYLKQVRODLQONPNSKDEGADILKKNSAL
ATLOAELDONLILKROSOELYNKKILFVNEQLQIMVNEPDHETTELISQIKTLOE
LNIWMANLQONLEROVNHESQLRKDFIAFKKALKSMENLTKNHRLLDQIATLOE
SEKLEKIMIDROAEYSEONISEINENIKOLELANNPLISKSLSQSDLEHLQONQE
NLEKMAKOEKFEYNDYNTVEKELLSRRLSNIIEQGTMRCAVAYMEONLEPNI
DYENGVTGGLSEHVYKFNRYIPLKVSDEKFTOEYSYHDMCLNOKKNPNLSIST
TPHGSLSRESLILFAEKDITIKQYVITLQFVLSDDERSDMLDYSINDKDSIKLK
FEKHSISLDSKLVITENGLEDPLNFCSCDEHPNIPHSQKMIITVOFPPRDSKDGND
PVVDVFIETLNNKSIQEDKSIKFKESCETPIALVLKILSDTKSFLLNLNDSKN
VKKLITSEVOTOLCKRKKLT"

CDS
complement(4748..5266)
/codon_start=1
/product="putative protein"
/protein_id="CAA91592.1"
/db_xref="gi:1061237"
/db_xref="SPTREMBL:Q12184"

CDS

/translation="MLKIVTRAGHTARISNIAHLRTSPSLTRTTTTRTLRPLPST
SFLNHGILKPKPEELKIFILIKDSQRTYECCEETILLIDAGHNLDMEACAGSC
ACSTCHVYDPPYDALPEPEDEMDLALGLTETSRIGQIKMSKDIDIRVALP
OMRNVNNDPS"
5010..5312

CDS

/codon_start=1
/product="putative protein"
/protein_id="CA91593.1"
/db_xref="GI:1061238"
/db_xref="SPTREMBL:Q12058"
/translation="MSRWVSPQTSYVFWPEPSFRIRKIVFSSSPGFPMKPKMEDY
EKGRNLVYVVRYSRDEVRKCAAFDILAVCPARVTFISGGGLSDITSAYV"
complement(5694..6104)

CDS

/codon_start=1
/product="putative protein"
/protein_id="CA91594.1"
/db_xref="GI:1061239"
/db_xref="SPTREMBL:Q12048"
/translation="MSSVSEPSISREDEVPFLSFSNVAEPFPMLSHSDSTYADP
SOSYQOELLFPLSYIVDESECTHTKDKTSDNITSTEDCOMFDMERNGAASAVAA
ASKESNASGAFASNDAPANVAQNYRLMSSV"
complement(8054..10738)

CDS

/codon_start=1
/product="putative protein"
/protein_id="CA91595.1"
/db_xref="GI:1061240"
/db_xref="SPTREMBL:Q12344"
/translation="MSSDKSTKNTDTLASFEVHGDNHSNNGSMEERIKSTPDQYE
EIAIPIETPLHSDKELNDKOSIGHEATPNYSREPIISGDEDTQITEOVANQORE
TREPSEIDNPLDPEKDTTDOVAPNGNIEKDYDKENERVADTAKVSPSPN
RYVTKNSNGSGESSKFLDDSGTTTANANDISISEVTPESSDNNQOIHITVE
VAGINLENKQKAAIEDGVTAENLSEETARKVPIPTQIINEKDNSENVSAI
PTSSSPPLPRONVATSTPKLPRGKQREOPKTKNAVPEPLEEEMKSEKFRNPE
TRNSYHVPPLGSKTAQLESAEINLJASRRTSHHINKEGEETRESIOGQSPFK
STPTEFLENSEYENVENVEDREMEFIDMSFMTVOYNDYATVASNEPERLEHVTN
GIPOIRGIIMOLMANSKREMEDIEYETILLDTCEHNTITRDLRKTYAEDMESEL
YKIVYVYDPOVGYTQMGFIAPPLINENAESGELVGLKNYGLRELFEP
PGLMILYQFORLLEHSPSLKRLIRIGISITWATQWELTFPAVKFPEFLFED
LYVFGIEVLKFAVNLMLKNEETLVKLRDELDFKDELFNLYLMGDDDAVYM
GLSNGNSKRGNDGTFSYVNDLFVDMATGVYITPLRLRYRGEYIEHKEOKEDH
YESLIQHOLQOREAKLEHYSILINKENISANELLIONRLNEMMLDEKNDLINTT
DIKSIIEEIKRONLPNDASLPADIREDEERTISRNNEMENGLERTTELQAE
IDELINIEKQVSTASILERDSKAKGKMGTFKVKY"
complement(11096..13741)

CDS

/codon_start=1
/product="GAL4"
/protein_id="CA91596.1"
/db_xref="GI:1061241"
/db_xref="SWISS-PROT:P04386"
/translation="MKLLSIEOACDICKLKIKCKEPRKCAKCKAKNNWEKYSKT
KRSPLTRAHLEVESRLERLEOFLIFRREDMDITIKMSIDIKALTLGLFVQDNV
NKDAVTRLAIVETDMLTLRQHRISATSSSESSNKGQQLTVSIDSAAHNDSTIP
LDPMRDLHGFDEEDDMSDGLPLKTPDNNNGFGDGLICILRISIGFRENTN
SVNRPLPTMTDITRTYLRSTSLRLOSILNHPYCPIVHSPTLMMLKLNNOIEITK
DOWOILFNCLITAGMCIEGSDIDVFOYONAKSHLTSVPESSGIIITVTLHLIS
YTOROKNTSYNHSFISIMASISLGNRLDPSGSSSILTBORRITMSYSWELQ
SLYGRSIQISQNTISPPSSVDVQRTTTPTYHGIETARLQVTFITLEDKVT
AESGPICAKCIMEIEVSRQAPREFLOMOTSTALTLWLEKHEWISFTRELKMK
OLSLITVLDFEPTNTQKKSQLEODNDHOSYEVRCISLMSDAQRYMVSYSYMD
NHNVTYFRANKSTYLFNAVLPIKTLSSKSNANENETAOLOQINIVMLKLKLA
TKIQTQEKTYOLEVECAPPLISQCAIPLPHISYNNSSGSAIKNIGSATYAOPTL
PEENVNNIISKTVYEGSPSPVPLKSGASFOLVYKLSNRPSSNPTITRSPSH
RSVPLFGQOQLOSLVPLTPSALFGGANOGSGADSSISPTFNSNGCNLTTO
TNSQALSGPIASSVHNDPMNNEITASKIDGNSKRLSPGWTDTATNAFSLITGMF
NTTMDVYNYLVDDDEDTTPNPKKE"
complement(14010..15581)

CDS

/codon_start=1
/product="putative protein"
/protein_id="CA91598.1"
/db_xref="GI:1061243"
/db_xref="SPTREMBL:Q12270"
/translation="NMKSYVPCGHPRAALITGLVLELTAIYLSIFLALREDLSLA
PESLFLQMSRLSLYPLHLSLPHLEFNLAIAWALNIEETHGTVYGVPLNLSALP
AGLYLCKRLKLPALVAGASGWCFTLEAYYSFESQIRPTRIFRDYISIPILYLP
LVLLVAIVAVIGSSFWGHFFGLCVGAYVIGYKESFENKITPPGMITFKIEKSLDGLR
LIPWGIKRYTRDDIDRTKDYEPIMSTETPLPHNDNSGTVIGTA"
16971..18335

CDS

/codon_start=1
/product="putative protein"
/protein_id="CA91599.1"
/db_xref="GI:1061244"
/db_xref="SPTREMBL:Q12179"
/translation="MHGPTSKAISRNVSRYKRRAPRPVYVTOAMNKLAVNTLSAO
EKLRREVLSFNRSLNLSOYKSDMKHPAMVIOGDGATSVLNSLPEIKOSFPS
SIDLHGHTNYLVNHPMLKLYIRISPFYIKSSISERTSLINNQKRLVADV
IYDEANHLATLSDAKRFRYGENHLKDLMSLCKVLVLYIDQALRMSYVDEGSNNA
TKDEYNTEPRPSRPWTLKQOFRAAONVNLNIDQISVAGKIPPIESVLSKGNDC
ADDKTKNDFKIMDCGAMYEAIKEKRDQYQCMRLSTYEPFLDGDYVDEKNDG
KTRMRYTPREVTWSEKRDIDEVGYTIOGDLNVAIVYILGRSISGYAANDCIL
RELDDNAGEFKKKKNHABDVOKIIMNSINVILTRGVGLVYVAYDELBRELLR
PSK"
complement(18399..19418)

CDS

/codon_start=1
/product="putative protein"
/protein_id="CA91600.1"
/db_xref="GI:1061245"
/db_xref="SPTREMBL:Q12520"
/translation="MAGSTSLVICAIGIYATFLMALVOEPLATRPWNSMCKPOP
NVLISIQASVAMMGVLYLNKRYEPRKRIKMKOLMISFQSSSGLPLATTSKL
HVDYLVLMKASKMIPYLVHLLYTPPLASOKYVALLVSGVITFTGGDNGKL
KRFNESGNDKSGFGLPSSLELDLITLAKMDKANEKQVTLTGALHMF
TLNLFVITMNLITYVDCQKMDNVAIVLMDPQWGYLMLVFCGAMGCEFTFYLE
OPGSLVIMITVTKKMSLITLIVFGKSVRFQWGMFTVFGITWEALNKKANIP
KAKSA"
19902..21701

CDS

/codon_start=1
/product="SRP68"
/protein_id="CA91601.1"
/db_xref="GI:1061246"
/db_xref="SWISS-PROT:P38687"
/translation="MVAVSPILITAGRAEQPLETSDSPAKYHAKLKLQHLRSCH
LVTDITKYSKKNKYGEINSEDDNKTGLIGVILLHAERDLALATLKLRLROGKL
KSEBEKLYSKKACGTADKLAVNVOENQOITTRAOYAFAPKIVSEYLINGRKK
KDNKISNNILAYFAALEHKNLSILAEVYDITVYKQYSLOVAGNLTTPENF
IVEROSDEKDEDELVYLILNENPNKKTITTEBDQVITINMRSFNKKIIDAQVQ
FLEGESLHPQITQYTORLSKMLADHREFTIAHDDODDIDEMEVSSNNQIIL
AYIRYNLILSISERDLPFLHNNOMLKNTSIPSLTKYKEMERVYALVYLSIDM
ELPEVYDLDLSLQDCKLYDFDLNTGCLSVLYOSKRYKEMERVYALVYLSIDM
LSEBESIDEILPANIILSNLSVLSKRLIENGNSVITLAEYERKNHGSGLKQVLT
IEKLSKKIIPITQIKNLPKMLPIPSKPTLPDLAFNITTDKQPSASQVKS

Query Match 85.5%: Score 18.8; DB 8; Length 37808;
Best Local Similarity 90.9%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACAA 22
DB 4025 AAGAAAAATCTAGACAA 4004

/translation="NDPFGNKRSGISYSGSSOROPYKKNNTLSTGNSPSAADOGR
GSPFGMSGNTTNGNSKNSGCDLSATTYASRPSYPLDPSPPVFSFNHSQLQAR

Tue Mar 18 16:16:15 2003

us-09-836-439-4.rge

Page 18

Search completed: March 17, 2003, 11:31:31
Job time : 548.495 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:42:12 ; Search time 126.086 Seconds
(without alignments)
392.938 Million cell updates/sec

Title: US-09-836-439-4

Perfect score: 22
Sequence: 1 aagagaaatctagacagca 22

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	80.9	2483	24	ABO69157
2	17.4	79.1	370	22	ABAI12703
3	17.4	79.1	412	22	ABAI1335
4	17.4	79.1	412	22	ABAI1336
5	17.2	78.2	434	23	ABV37860
6	17.2	78.2	507	23	ABV38334
7	17.2	78.2	674	22	AAS30250
8	17.2	78.2	674	22	AAS30251
9	17.2	78.2	674	22	AAS26974

10	17.2	78.2	674	22	AAS26975	Human genomic DNA
11	17.2	78.2	674	22	AAS33485	DNA encoding human
12	17.2	78.2	674	22	AAS33486	DNA encoding human
13	17.2	78.2	782	24	AAD37775	Extended sequence
14	17.2	78.2	786	24	ABQ73702	Human colon specific
15	17.2	78.2	831	24	ABQ43712	Oligonucleotide fo
16	17.2	78.2	831	24	ABQ43713	Oligonucleotide fo
17	17.2	78.2	1372	23	ABV25096	Human prostate exp
18	17.2	78.2	1509	24	ABQ73703	Human colon specif
19	17.2	78.2	1588	21	AAC35087	Arabidopsis thalia
20	17.2	78.2	1622	22	AAK83087	Human immune/haema
21	17.2	78.2	2713	23	ABL24452	Drosophila melanog
22	17.2	78.2	2961	24	ABQ77508	Human cytokine rec
23	17.2	78.2	2963	24	ABQ77508	Drosophila melanog
24	17.2	78.2	3138	23	ABL24782	Human wild-type fc
25	17.2	78.2	3729	22	AAE77688	Drosophila melanog
26	17.2	78.2	4081	23	ABL16178	Human variant fcp
27	17.2	78.2	4131	22	AAE77689	Human immune/haema
28	17.2	78.2	4258	22	AAK67062	Human immune/haema
29	17.2	78.2	5552	21	AAAS58309	Human immune/haema
30	17.2	78.2	6716	24	ABL33783	PIP/PyCSP.1 plasm
31	17.2	78.2	7202	22	ABAI19579	Human immune syst
32	17.2	78.2	7202	22	ABAI19579	Human immune syst
33	17.2	78.2	9615	22	AAI36590	Human immune/haema
34	17.2	78.2	11298	18	AAE86756	Human high affinity
35	17.2	78.2	11298	19	AAV54661	Human beta subunit
36	17.2	78.2	11298	21	AAE720937	Human high affinity
37	17.2	78.2	11298	21	AAA34815	Human adenosine re
38	17.2	78.2	11298	22	AAE92144	Human FCGR1 beta c
39	17.2	78.2	11357	14	AAO51024	Human FCGR1 beta c
40	17.2	78.2	21742	21	AAE20938	Human high affinity
41	17.2	78.2	21742	21	AAE20938	Human high affinity
42	17.2	78.2	24512	22	ABAI19578	Human adenosine re
43	17.2	78.2	24512	22	AAK67571	Human immune/haema
44	17.2	78.2	24512	22	AAK71749	Human immune/haema
45	17.2	78.2	33030	22	AAE29337	Atopy related gene

ALIGNMENTS

RESULT 1	ABO69157	standard; DNA: 2483 BP.
ID	ABO69157	
AC	ABO69157	
XX	29-AUG-2002 (first entry)	
DE	Listeria monocytogenes 4b contig DNA sequence #1923.	
XX	Antibacterial; Listeria; food contamination; mutational analysis;	
KW	Infection; ds.	
XX	Listeria monocytogenes 4b.	
OS	WO200228891-A2.	
XX	11-APR-2002.	
PD	04-OCT-2001; 2001WO-FR03061.	
XX	04-OCT-2000; 2000FR-0012697.	
PR	(INSP) INST PASTEUR.	
PA	(CNRS) CNRS CENT NAT RECH SCI.	
XX	Kunst F, Glaser P;	
PI	WPI; 2002-332479/37.	
XX	New genomic sequences from Listeria species, useful for detection,	
DR	Treatment and prevention of infection, also related polypeptides,	
XX		
PT		

PT antibodies and modulators -
XX
PS Claim 14; SEQ ID 1970; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes
CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIP0 at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2483 BP; 771 A; 651 C; 362 G; 699 T; 0 other;

Query Match 80.9%; Score 17.8; DB 24; Length 2483;
Best Local Similarity 90.5%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTGACACGCA 21
|||||
DB 2200 AAGAAAAATCTGACACGCA 2220

RESULT 2
ID ABA12703/C
ID ABA12703 standard; cDNA; 370 BP.
XX ABA12703;
XX
DT 23-JAN-2002 (First entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 1710.
XX
KW Human; noctropic; neuroprotective; cytosolic; dermatological; virologic;
KW immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulvular;
KW antiparkinsonian; antischizophrenic; antidiabetic; antitubercular; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antineoplastic;
KW allergic; antidiabetic; antileukemic; anticonvulsant; antitubercular;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 18-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 22-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234274.
PR 21-SEP-2000; 2000US-0234275.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

08-NOV-2000: 2000US-0246478.
PR 08-NOV-2000: 2000US-0246523.
PR 08-NOV-2000: 2000US-0246524.
PR 08-NOV-2000: 2000US-0246525.
PR 08-NOV-2000: 2000US-0246526.
PR 08-NOV-2000: 2000US-0246527.
PR 08-NOV-2000: 2000US-0246528.
PR 08-NOV-2000: 2000US-0246532.
PR 08-NOV-2000: 2000US-0246609.
PR 08-NOV-2000: 2000US-0246610.
PR 08-NOV-2000: 2000US-0246611.
PR 08-NOV-2000: 2000US-0246613.
PR 17-NOV-2000: 2000US-0249207.
PR 17-NOV-2000: 2000US-0249208.
PR 17-NOV-2000: 2000US-0249209.
PR 17-NOV-2000: 2000US-0249210.
PR 17-NOV-2000: 2000US-0249211.
PR 17-NOV-2000: 2000US-0249212.
PR 17-NOV-2000: 2000US-0249213.
PR 17-NOV-2000: 2000US-0249214.
PR 17-NOV-2000: 2000US-0249215.
PR 17-NOV-2000: 2000US-0249216.
PR 17-NOV-2000: 2000US-0249217.
PR 17-NOV-2000: 2000US-0249218.
PR 17-NOV-2000: 2000US-0249244.
PR 17-NOV-2000: 2000US-0249245.
PR 17-NOV-2000: 2000US-0249264.
PR 17-NOV-2000: 2000US-0249265.
PR 17-NOV-2000: 2000US-0249297.
PR 17-NOV-2000: 2000US-0249299.
PR 17-NOV-2000: 2000US-0249300.
PR 01-DEC-2000: 2000US-0250391.
PR 01-DEC-2000: 2000US-0251160.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251030.
PR 05-DEC-2000: 2000US-0251988.
PR 06-DEC-2000: 2000US-0251479.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251856.
PR 08-DEC-2000: 2000US-0251859.
PR 08-DEC-2000: 2000US-0251869.
PR 08-DEC-2000: 2000US-0251889.
PR 11-DEC-2000: 2000US-0251990.
PR 11-DEC-2000: 2000US-0254097.
PR 05-JAN-2001: 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-541565/60.
P-PSDB: ABA16377.

Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -

PS Claim 1; SEQ ID NO 1710; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABA1678-ABA16001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus; Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 370 BP; 130 A; 59 C; 60 G; 120 T; 1 other:

Query Match 79.1%; Score 17.4; DB 22: Length 370;
Best Local Similarity 94.7%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACAG 19
DB 223 AAGAAAAATCTAGACAG 205
||||| |||||||
ABAB1835/c
ID ABA18135 standard; DNA; 412 BP.
XX
AC ABA18135;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 10466.
XX
XX Human; noctropic; neuroprotective; cytostatic; dermatological; virocidic;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisticking; antianaemic; antitubercitic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
OS
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
XX 17-JAN-2001: 2001WO-US01334.
PF
XX 31-JAN-2000: 2000US-0179065.
XX 04-FEB-2000: 2000US-0180628.
PR 24-FEB-2000: 2000US-0184664.
PR 02-MAR-2000: 2000US-0186350.
PR 16-MAR-2000: 2000US-0189874.
PR 17-MAR-2000: 2000US-0190076.
PR 18-APR-2000: 2000US-0198123.
PR 19-MAY-2000: 2000US-0205515.
PR 07-JUN-2000: 2000US-0209467.
PR 28-JUN-2000: 2000US-0214886.
PR 30-JUN-2000: 2000US-0215135.
PR 07-JUL-2000: 2000US-0216647.
PR 07-JUL-2000: 2000US-0216880.
PR 11-JUL-2000: 2000US-0217487.
PR 11-JUL-2000: 2000US-0217496.
PR 14-JUL-2000: 2000US-0218290.
PR 26-JUL-2000: 2000US-0220963.
PR 26-JUL-2000: 2000US-0220964.
PR 14-AUG-2000: 2000US-0224518.
PR 14-AUG-2000: 2000US-0224519.
PR 14-AUG-2000: 2000US-0225213.
PR 14-AUG-2000: 2000US-0225214.
PR 14-AUG-2000: 2000US-0225214.
PR 14-AUG-2000: 2000US-0225266.
PR 14-AUG-2000: 2000US-0225267.
PR 14-AUG-2000: 2000US-0225268.
PR 14-AUG-2000: 2000US-0225270.
PR 14-AUG-2000: 2000US-0225447.
PR 14-AUG-2000: 2000US-0225757.
PR 14-AUG-2000: 2000US-0225758.
PR 14-AUG-2000: 2000US-0225759.
PR 18-AUG-2000: 2000US-0226279.
PR 22-AUG-2000: 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0228287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0233081.
 PR 14-SEP-2000; 2000US-0233397.
 PR 14-SEP-2000; 2000US-0233398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 25-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-541565/60.
 DR Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Disclosure; SEQ ID NO 10466; 1701bp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (AB14678-AB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 412 BP; 137 A; 69 C; 71 G; 135 T; 0 other;
 Query Match 79.1%; Score 17.4; DB 22; Length 412;
 Best Local Similarity 94.7%; Pred. No. 3,8e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAGAAAAATCTAGACAG 19
 DB 220 AAGAAAAATCTAGACAG 202

RESULT 4
ABAI8136/C
ID ABAI8136 standard; DNA: 412 BP.
XX
AC ABAI8136;
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 10467.
XX
KW Human; noctropic; neuroprotective; cytoskeletal; dermatological; virocidic;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antistickling; antianaemic; antiahrthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN MO200159063-A2.
XX
PD 16-AUG-2001.
XX
FE 17-JAN-2001; 2001MO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239393.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.

XX	17-NOV-2000;	2000US-0249265.		
PR	17-NOV-2000;	2000US-0249297.		
PR	17-NOV-2000;	2000US-0249299.		
PR	17-NOV-2000;	2000US-0249300.		
PR	01-DEC-2000;	2000US-0250391.		
PR	01-DEC-2000;	2000US-0251160.		
PR	05-DEC-2000;	2000US-0251988.		
PR	05-DEC-2000;	2000US-0256719.		
PR	06-DEC-2000;	2000US-0251479.		
PR	08-DEC-2000;	2000US-0251856.		
PR	08-DEC-2000;	2000US-0251869.		
PR	08-DEC-2000;	2000US-0251989.		
PR	08-DEC-2000;	2000US-0251990.		
PR	11-DEC-2000;	2000US-0254097.		
PR	05-JAN-2001;	2001US-0259678.		
XX	(HUMA-) HUMAN GENOME SCI INC.			
PA	Rosen CA, Barash SC, Ruben SM;			
PI	WPI; 2001-541565/60.			
DR	Nucleic acids encoding 3224 human nervous system antigen polypeptides,			
PT	useful for preventing, diagnosing and/or treating nervous system			
PT	cancers and metastases -			
XX	Disclosure; SEQ ID NO 10467; 1701pp + Sequence listing; English.			
PS	The invention relates to novel genes (ABAI1004-ABA21534) and proteins			
CC	(ABAI4678-ABH18001) useful for preventing, treating or ameliorating			
CC	medical conditions e.g. by protein or gene therapy. The genes are			
CC	isolated from a range of human tissues disclosed in the specification.			
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful			
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast			
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone			
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;			
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune			
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's			
CC	disease, multiple sclerosis, Rheumatoid arthritis and ulcerative			
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;			
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and			
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal			
CC	and parasitic infections.			
CC	Note: The sequence data for this patent did not form part of the			
CC	printed specification, but was obtained in electronic format directly			
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.			
XX	Sequence 412 BP; 137 A; 69 C; 71 G; 135 T; 0 other;			
SO				
Query Match				
Best Local Similarity 79.1%; Score 17.4; DB 22; Length 412;				
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
OY	1 AAGAAATAATCTAGACAAG 19			
DB				
220 AAGAAATATCTAGACAAG 202				
RESULT 5				
ABV37860/C				
ID	ABV37860 standard; cDNA; 434 BP.			
XX	ABV37860;			
AC	16-SEP-2002 (first entry)			
DT	Human prostate expression marker CDNA 37851.			
XX	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;			
DE	pharmacogenomic marker; gene; ss.			
XX				

XX	Homo sapiens.
XX	
PN	WO200160860-A2.
XX	
PD	23-AUG-2001.
XX	
PF	20-FEB-2001; 2001WO-US05171.
XX	
PR	17-FEB-2000; 2000US-183319P.
PR	16-MAR-2000; 2000US-189862P.
PR	25-MAY-2000; 2000US-207454P.
PR	09-JUN-2000; 2000US-211314P.
PR	18-JUL-2000; 2000US-218007P.
PR	13-DEC-2000; 2000US-255281P.
XX	
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	
XI	Schlegel R, Endege WO, Monahan JE;
DR	WPI; 2001-662795/76.
XX	
PT	Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -
PT	
PS	Claim 1; Page 7745-7746; 11750pp; English.
XX	
CC	The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:
CC	(a) assessing whether a patient is afflicted with prostate cancer;
CC	(b) monitoring the progression of prostate cancer in a patient;
CC	(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
CC	(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;
CC	(f) assessing the prostate cell carcinogenic potential of a compound;
CC	(g) determining whether prostate cancer has metastasized in a patient;
CC	(h) assessing the aggressiveness or indolence of prostate cancer in a patient;
CC	(I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX	
SQ	Sequence 434 BP; 130 A; 78 C; 76 G; 150 T; 0 other;
	Query Match 78.2%; Score 17.2; DB 23; Length 434;
	Best Local Similarity 86.4%; Pred. No. 4.6e+02;
	Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0.
OY	1 AGAAGAAAATCTAGACAAACGCAA 22 DB 240 AAGAAGAAAATCTGAAGAACGAAA 219
RESULT 6	
ABV58334/c	
ID	ABV58334 standard; CDNA; 507 BP.
XX	
AC	ABV58334;
XX	
DT	13-SEP-2002 (first entry)
XX	
DE	Human prostate expression marker CDNA 58325.
XX	
KX	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KV	pharmacogenomic marker; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200160860-A2.
XX	
PD	23-AUG-2001.
XX	

PF 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183119P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX WPI: 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1: Page 11199; 11750P; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABY00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 507 BP; 179 A; 98 C; 93 G; 134 T; 3 other:
SO
Query Match 78.2%; Score 17.2; DB 23; Length 507;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAGAAAAATCTAGACAAGCA 22
Db 296 AAGAAAAACACTACTCAAGCA 275
RESULT 7
AAS30250
ID AAS30250 standard; DNA: 674 BP.
XX
XX AAS30250:
XX
XX 21-NOV-2001 (first entry)
XX
XX DNA encoding renal and cardiovascular-associated protein, Seq ID 168.
XX
XX Human; antiinflammatory; neuroprotective; immunomodulator; vulnery;
KW cardiovascular; cytostatic; nephrotoxic; antineoplastic; nephritis;
KW immunosuppressive; kidney disorder; renal failure; hypertension;
KW cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
KW blood coagulation disorder; electrolyte imbalance disorder; cancer;
KW hypotension; hyperkalemia; neoplastic disorder; nephroma;
KW autoimmune disease; inflammatory disease; reproductive system disorder;
KW endocrine disorder; neural activity; neurological disorder;
KW wound healing; respiratory disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200155328-A2.
XX
XX 02-AUG-2001.
PD

XX
XX 17-JAN-2001; 2001WO-US01359.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216680.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217486.
XX 14-JUL-2000; 2000US-0218290.
XX 16-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225427.
XX 14-AUG-2000; 2000US-0225477.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234974.
XX 25-SEP-2000; 2000US-0234998.
XX 26-SEP-2000; 2000US-0235484.
XX 27-SEP-2000; 2000US-0235834.
XX 27-SEP-2000; 2000US-0235836.
XX 29-SEP-2000; 2000US-0236327.
XX 29-SEP-2000; 2000US-0236367.
XX 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 06-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259618.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-488787/53.
 DR XX

PR New polynucleotides and polypeptides, useful for diagnosing, treating,
 PT preventing or prognosing e.g. kidney, cardiovascular, blood,
 PT electrolyte imbalance or neoplastic disorders, autoimmune diseases,
 PT cancers -
 XX
 PS Claim 1; SEQ ID No 168; 506bp; English.
 XX
 CC The invention relates to novel nucleic acids and polypeptides useful for
 CC diagnosing, treating, preventing and/or prognosing disorders related to
 CC these polypeptides. The polynucleotides are especially useful in the
 CC diagnosis, prognosis, prevention and/or treatment of diseases which
 CC include kidney disorders (e.g. renal failure or nephritis),
 CC cardiovascular disorders (e.g. hypertension or myocardial infarction),
 CC blood disorders (e.g. anaemia or blood coagulation disorders),
 CC electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),
 CC neoplastic disorders (e.g. nephroma or renal cell cancer), autoimmune
 CC diseases, cancers, inflammatory diseases, reproductive system
 CC disorders, endocrine disorders, neural activity and neurological
 CC disorders, wound healing and respiratory disorders. AAS30251
 CC represent the novel human renal and cardiovascular-associated nucleic
 CC acid sequences of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at:
 CC ftp.wipo.int/pub/published_pat_sequences.
 CC
 SQ Sequence 674 BP; 144 A; 184 C; 145 G; 201 T; 0 other;
 XX
 Query Match 78.2%; Score 17.2; DB 22; Length 674;
 Best Local Similarity 86.4%; Pred. No. 4.6e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 MAGAAAAAATCTAGACAGCAA 22
 Db 137 MAGAAAAAATTTAACAACAA 158
 |||||
 AAS30251
 AAS30251 standard; DNA; 674 BP.
 XX
 AC AAS30251;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE DNA encoding renal and cardiovascular-associated protein, Seq ID 169.
 XX
 KW Human; antiinflammatory; neuroprotective; immunomodulator; vulnerary;
 KW cardiovascular; cytosolic; nephrotropic; antianemic; nephritis;
 KW immunosuppressive; kidney disorder; renal failure; hypertension;
 KW cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
 KW blood coagulation disorder; electrolyte imbalance disorder; cancer;
 KW hypotonaemia; hyperkalaemia; neoplastic disorder; nephroma;
 KW autoimmune disease; inflammatory disease; reproductive system disorder;
 KW endocrine disorder; neural activity; neurological disorder;
 KW wound healing; respiratory disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155328-A2.
 XX
 PD 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001MO-US01359.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0215647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226886.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236379.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239835.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488787/53.

New polynucleotides and polypeptides, useful for diagnosing, treating, preventing or prognosing e.g. kidney, cardiovascular, blood, electrolyte imbalance or neoplastic disorders, autoimmune diseases, cancers -
Claim 1; SEQ ID No 169; 506bp; English.

The invention relates to novel nucleic acids and polypeptides useful for diagnosing, treating, preventing and/or prognosing disorders related to these polypeptides. The polynucleotides are especially useful in the diagnosis, prognosis, prevention and/or treatment of diseases which include kidney disorders (e.g. renal failure or nephritis),

CC cardiovascular disorders (e.g. hypertension or myocardial infarction),
CC blood disorders (e.g. anaemia or blood coagulation disorders),
CC electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),
CC neoplastic disorders (e.g. nephroma or renal cell cancer), autoimmune
CC diseases, cancers, inflammatory diseases, reproductive system
CC disorders, endocrine disorders, neural activity and neurological
CC disorders, wound healing and respiratory disorders. AAS30165-AAS30251
CC represent the novel human renal and cardiovascular-associated nucleic
CC acid sequences of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIP0 at:
CC ftp.wipo.int/pub/published_pct_sequences.
CC
CC
SQ Sequence 674 BP; 144 A; 184 C; 145 G; 201 T; 0 other;
Query Match 78.2%; Score 17.2; DB 22; Length 674;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 AAGAAAAATCTGACAAAGCA 22
|||||||
DB 137 AAGAAAAATTTAAACAAACA 158
RESULT 9
AAS26974
ID AAS26974 standard; DNA; 674 BP.
AC AAS26974;
XX
XX 07-NOV-2001 (first entry)
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 310.
XX
XX Human; immunosuppressive; antiarthritic; ds; antirheumatic;
KW cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnery; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
OS Homo sapiens.
XX
XX WO20015441-A2.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01320.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0218860.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239835.
PR 13-OCT-2000; 2000US-0239837.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241885.
PR 20-OCT-2000; 2000US-0241886.
PR 20-OCT-2000; 2000US-0241887.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

CC	08-NOV-2000	2000US-0246525	PR
PR	08-NOV-2000	2000US-0246525	PR
PR	08-NOV-2000	2000US-0246526	PR
PR	08-NOV-2000	2000US-0246527	PR
PR	08-NOV-2000	2000US-0246528	PR
PR	08-NOV-2000	2000US-0246532	PR
PR	08-NOV-2000	2000US-0246539	PR
PR	08-NOV-2000	2000US-0246610	PR
PR	08-NOV-2000	2000US-0246611	PR
PR	08-NOV-2000	2000US-0246613	PR
PR	17-NOV-2000	2000US-0249214	PR
PR	17-NOV-2000	2000US-0249215	PR
PR	17-NOV-2000	2000US-0249216	PR
PR	17-NOV-2000	2000US-0249217	PR
PR	17-NOV-2000	2000US-0249218	PR
PR	17-NOV-2000	2000US-0249244	PR
PR	17-NOV-2000	2000US-0249245	PR
PR	17-NOV-2000	2000US-0249264	PR
PR	17-NOV-2000	2000US-0249265	PR
PR	17-NOV-2000	2000US-0249297	PR
PR	17-NOV-2000	2000US-0249300	PR
PR	01-DEC-2000	2000US-0250150	PR
PR	01-DEC-2000	2000US-0250391	PR
PR	05-DEC-2000	2000US-0251030	PR
PR	05-DEC-2000	2000US-0251988	PR
PR	05-DEC-2000	2000US-0256719	PR
PR	08-DEC-2000	2000US-0251479	PR
PR	08-DEC-2000	2000US-0251856	PR
PR	08-DEC-2000	2000US-0251866	PR
PR	08-DEC-2000	2000US-0251989	PR
PR	11-DEC-2000	2000US-0254190	PR
PR	05-JAN-2001	2001US-0254076	PR
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-476222/51.		
XX			
PT	Novel polypeptides and polynucleotides useful as diagnostic reagents to		
PT	diagnose diseases or disorders associated with aberrant expression or		
PT	activity of polypeptides, for treating blood clotting disorder,		
XX	hemophilia		
XX			
XX	Disclosure; SEQ ID NO 310; 601pp; English.		
XX			
CC	The invention relates to isolated nucleic acid molecules and their		
CC	encoded secreted proteins. The nucleic acids and proteins are used to		
CC	prevent, treat or ameliorate a medical condition in e.g. humans, mice,		
CC	rabbits, goats, horses, cats, dogs, chickens or sheep. They		
CC	are also used in diagnosing a pathological condition or susceptibility		
CC	to a pathological condition. Antibodies to the proteins can also		
CC	be used in alleviating symptoms associated with the disorders and in		
CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked		
CC	immunosorbent assays (ELISA). Disorders which are diagnosed or treated		
CC	include autoimmune diseases e.g. Rheumatoid arthritis,		
CC	hyperproliferative disorders e.g. neoplasms of the breast or liver,		
CC	cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders		
CC	e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.		
CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi		
CC	and ocular disorders e.g. corneal infection, and many other		
CC	disorders listed in the specification. The polypeptides can also		
CC	be used to aid wound healing and epithelial cell proliferation, to		

CC	prevent skin aging due to sunburn, to maintain organs before
CC	transplantation, for supporting cell culture of primary tissues, to
CC	regenerate tissues and in chemotaxis. The polypeptides can also be used
CC	as a food additive or preservative to increase or decrease storage
CC	capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC	minerals, cofactors and other nutritional components. The present
CC	sequence is a genomic DNA encoding a partial novel secreted protein of
CC	
Query Match	78.2%: Score 17.2: DB 22: Length 674:
Best Local Similarity	86.4%: Pred. No. 4, 6e+02:
Matches 19: Conservative	0: Mismatches 3: Indels 0: Gaps 0
OY	1 AAGAAAAATCTGACAAACGAA 22
DB	137 AAGAAAAATTTAAACAAACAA 158
RESULT 10	
AAS26975	
ID	AAS26975 standard; DNA; 674 BP.
XX	
AC	AAS26975;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human genomic DNA encoding partial novel secreted protein, Seq ID 311.
XX	
KW	Human; immunosuppressive; antiarthritic; ds; antirheumatic;
KW	cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW	neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW	vulnerable; secreted protein; rheumatoid arthritis;
KW	hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW	cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW	neurotic system disorder; Alzheimer's disease; infection; ocular disorder;
KW	corneal infection; wound healing; epithelial cell proliferation;
KW	skin ageing; food additive; preservative; antiproliferative.
XX	
OS	Homo sapiens.
XX	
PN	WO200155441-A2.
XX	
PD	02-AUG-2001.
XX	
XX	
PF	17-JAN-2001; 2001MO-US01320.
XX	
31-JAN-2000;	2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231442.
PR 08-SEP-2000; 2000US-0231443.
PR 08-SEP-2000; 2000US-0231444.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-476222/51.
XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
XX diagnose diseases or disorders associated with aberrant expression or
XX activity of polypeptides, for treating blood clotting disorder,
XX haemophilia
XX
XX Disclosure; SEQ ID No 311; 601pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
XX encoded secreted proteins. The nucleic acids and proteins are used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They
XX are also used in diagnosing a pathological condition or susceptibility
XX to a pathological condition. Antibodies to the proteins can also
XX be used in alleviating symptoms associated with the disorders and in
XX diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunoassay assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi
XX and ocular disorders e.g. corneal infection, and many other
XX disorders listed in the specification. The polypeptides can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used
XX as a food additive or preservative to increase or decrease storage
XX capabilities, fat content, lipid, protein, carbohydrate, vitamins,
XX minerals, cofactors and other nutritional components. The present
XX sequence is a genomic DNA encoding a partial novel secreted protein of

Query Match 78.2%; Score 17.2; DB 22; Length 674;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AAGAAAAAATCTAGACAGCAA 22
137 AAGAAAAAATTTAAACAAACAA 158

RESULT 11
AAS33485
ID AAS33485 standard; DNA; 674 BP.
XX AAS33485;
AC AAS33485;
XX 04-DEC-2001 (first entry)
DT
XX DNA encoding human secreted protein, Seq ID No 768.

Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
multiple sclerosis; cancer; hyperproliferative disorder; infection;
Gaucher's disease; neurological disease; cerebrovascular disorder;
thrombosis; wound healing; ds.

Homo sapiens.
XX
XX WO200155326-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01347.
XX 31-JAN-2000; 2000US-0179065.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451931/48.

New nucleic acids and polypeptides, useful for diagnosing, preventing
or treating medical conditions -
XX
XX Disclosure; SEQ ID No 768; 753bp; English.

The invention relates to novel isolated nucleic acid molecules (I)
encoding human secreted proteins (II). (I) and (II) are used to prevent,
treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
the prevention, treatment and diagnosis of diseases associated with
inappropriate expression of secreted proteins. (I) and complementary
sequences may also be used as DNA probes in diagnostic assays (e.g.
polymerase chain reactions (PCR)) to detect and quantitate the presence
of similar nucleic acid sequences in samples, and so which patients may
be in need of restorative therapy. (II) may also be used as antigens in
the production of antibodies and in assays to identify modulators
(agonists and antagonists) of the expression and activity of the secreted
proteins. The anti-(II) antibodies and antagonists may also be used to
down regulate expression and activity of (II). The anti-(II) antibodies
may also be used as diagnostic agents for detecting the presence of (II)
in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
disorders include for example: immune/autoimmune diseases (e.g. HIV
(human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
melanomas, neoplasms of the breast or liver, Sezary syndrome and
Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
cerebrovascular disorders (e.g. cardiac arrest, tachycardia,
angina and thrombosis), infections caused by bacteria, viruses and

CC fungi and ocular disorders (e.g. corneal infections). (I) and (II),
CC agonists, antagonists and antibodies can also be used to promote wound
CC healing, maintain organs before transplantation, and support cell culture
CC of primary tissues. AAS33043-AAS33486 represent human secreted protein
CC coding sequences, PCR primers, and related sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification but was obtained in electronic format directly from WIPO
CC at: ftp.wipo.int/pub/published_pcr_sequences.
XX
XX Sequence 674 BP; 144 A; 184 C; 145 G; 201 T; 0 other;

Query Match 78.2%; Score 17.2; DB 22; Length 674;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 AAGAAAAAATCTAGACAGCAA 22
137 AAGAAAAAATTTAAACAAACAA 158

RESULT 12
AAS33486
ID AAS33486 standard; DNA; 674 BP.
XX AAS33486;
AC AAS33486;
XX 04-DEC-2001 (first entry)
DT
XX DNA encoding human secreted protein, Seq ID No 769.

Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
multiple sclerosis; cancer; hyperproliferative disorder; infection;
Gaucher's disease; neurological disease; cerebrovascular disorder;
thrombosis; wound healing; ds.

Homo sapiens.
XX
XX WO200155326-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01347.
XX 31-JAN-2000; 2000US-0179065.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451931/48.

New nucleic acids and polypeptides, useful for diagnosing, preventing
or treating medical conditions -
XX
XX Disclosure; SEQ ID No 769; 753bp; English.

The invention relates to novel isolated nucleic acid molecules (I)
encoding human secreted proteins (II). (I) and (II) are used to prevent,
treat or ameliorate a medical condition in e.g. humans, mice, rabbits,
goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
the prevention, treatment and diagnosis of diseases associated with
inappropriate expression of secreted proteins. (I) and complementary
sequences may also be used as DNA probes in diagnostic assays (e.g.
polymerase chain reactions (PCR)) to detect and quantitate the presence
of similar nucleic acid sequences in samples, and so which patients may
be in need of restorative therapy. (II) may also be used as antigens in
the production of antibodies and in assays to identify modulators
(agonists and antagonists) of the expression and activity of the secreted
proteins. The anti-(II) antibodies and antagonists may also be used to
down regulate expression and activity of (II). The anti-(II) antibodies

CC may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anemias, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Parkinson's disease), neurological diseases (e.g. Alzheimer's disease, Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and thrombosis), infections caused by bacteria, viruses and fungi and ocular disorders (e.g. corneal infections). (I) and (II), agonists, antagonists and antibodies can also be used to promote wound healing, maintain organs before transplantation, and support cell culture of primary tissues. AAS3043-AAS3486 represent human secreted protein coding sequences, PCR primers, and related sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification but was obtained in electronic format directly from WIPO at: ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 674 BP; 144 A; 184 C; 145 G; 201 T; 0 other;

Query Match 78.2%; Score 17.2; DB 22; Length 674;
Best Local Similarity 86.4%; Pred. No. 4.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAGAAAAATCTAGACAGCAA 22
||| ||||| ||| ||| |||
DB 137 AAGAAAAATTTAAACAAACAA 158

RESULT 13

AAD37775/c
ID AAD37775 standard; DNA; 782 BP.

AC AAD37775;

DT 27-AUG-2002 (first entry)

XX Extended sequence for mouse IMX5_09.

DE Inflammatory bowel disease; IBD; autoimmune disorder; arthritis; allergy; haematopoietic cell; thrombolytic; blood coagulation disorder; nephritis; asthma; organ rejection; graft-versus-host disease; inflammation; shock; nerve disease; Alzheimer's disease; Parkinson's disease; antibacterial; Huntington's disease; immunosuppressive; sepsis; nephrotropic; nootropic; neuroprotective; anticonvulsant; gene therapy; mouse; ds.

XX Mus musculus.

OS WO200231116-A2.

PN 18-APR-2002.

PD 11-OCT-2001; 2001WO-US32176.

PF 11-OCT-2000; 2000US-239712P.

PR (DIGI-) DIGITAL GENE TECHNOLOGIES INC.

PA Viney JL, Sims JE, Dubose RF, Baum PR, Hasel KW, Hilbush BS;

PI WPI; 2002-426280/45.

DR New polynucleotide associated with inflammatory bowel disease for treating disorders of the immune system, nervous system, hematopoietic cells and to modulate inflammation

PS Claim 1; Page 202-203; 214pp; English.

CC The invention relates to an isolated polynucleotide associated with inflammatory bowel disease (IBD). The invention is useful for manufacturing a medicament for use in preventing, treating, modulating, or ameliorating a medical condition which is IBD. The polypeptide and

CC polynucleotide are useful for treating disorders of the immune system e.g. autoimmune disorders, deficiencies or disorders of haematopoietic cells, to modulate hemostatic, or thrombolytic activity, treat blood coagulation disorders, allergic reactions and conditions such as asthma, treat and/or prevent organ rejection or graft-versus-host disease and modulate inflammation, including inflammation associated with infection, shock, sepsis, arthritis and nephritis. The invention is useful to CC differentiate, proliferate and attract cells, leading to the regeneration of tissues and to treat central and peripheral nerve diseases e.g. Alzheimer's disease, Parkinson's disease, and Huntington's disease. The CC invention is useful in gene therapy. The present sequence is an extended sequence for mouse IMX5_09 which is used in the exemplification of the CC invention.

SO Sequence 782 BP; 212 A; 180 C; 138 G; 252 T; 0 other;

Query Match 78.2%; Score 17.2; DB 24; Length 782;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAGAAAAATCTAGACAGCAA 22
||| ||||| ||| ||| |||
DB 393 AAAAAAATCTAAACAAACAA 372

RESULT 14

ABQ73702/c
ID ABQ73702 standard; cDNA; 786 BP.

AC ABQ73702;

DT 07-OCT-2002 (first entry)

XX Human colon specific nucleic acid (CSNA) SEQ ID NO:8.

DE Human; colon specific nucleic acid; colon specific polypeptide; CSP; CSNA; colon specific gene; CSG; colon cancer; gene therapy; vaccine; cytosolic; gene; ss.

XX Homo sapiens.

OS WO200248370-A2.

PN 20-JUN-2002.

PD 30-OCT-2001; 2001WO-US51341.

PF 31-OCT-2000; 2000US-244717P.

PR (DIAD-) DIADEXUS INC.

PA Sun Y, Reclon H, Ghosh MG, Liu C;

PI WPI; 2002-583520/62.

DR Colon specific polypeptides and polynucleotides useful for detecting, diagnosing, monitoring, treating, staging and predicting cancers in humans having cancer and non-cancerous colon disease

PS Claim 1; Page 153; 243pp; English.

XX ABQ73695 to ABQ73841 represent human colon specific nucleic acid (CSNA) sequences, and ABP51826 to ABP51928 represent human colon specific polypeptide (CSP) sequences from the present invention. CSNA and CSP sequences have cytostatic activity, and can be used in gene therapy, antisense therapy and in vaccines. CSNA and CSP sequences can be used for diagnosing and monitoring the presence and metastases of colon cancer in a patient, by determining an amount of CSP or CSNA in a sample CC in a normal control, and comparing it to the amount of colon specific marker acid or the polypeptide in the sample compared to that of normal control CC is associated with presence of colon cancer. CSP and CSNA sequences can be used for producing engineered colon tissue for treatment and research.

CC CSNA sequences are useful for producing transgenic animals and cells
CC and also in gene therapy.

XX Sequence 786 BP; 256 A; 118 C; 140 G; 272 T; 0 other;

SO Query Match 78.2%; Score 17.2; DB 24; Length 786;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACACGAA 22
Db 715 ATGAGAAAAATCTAACAACGAA 694

RESULT 15
ABQ43712/c
ID ABQ43712 standard; DNA; 831 BP.

AC ABQ43712;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30303.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPICENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guettig D;

DR WPI: 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -

PS Claim 12: 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridized to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridization to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridized to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ43712-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

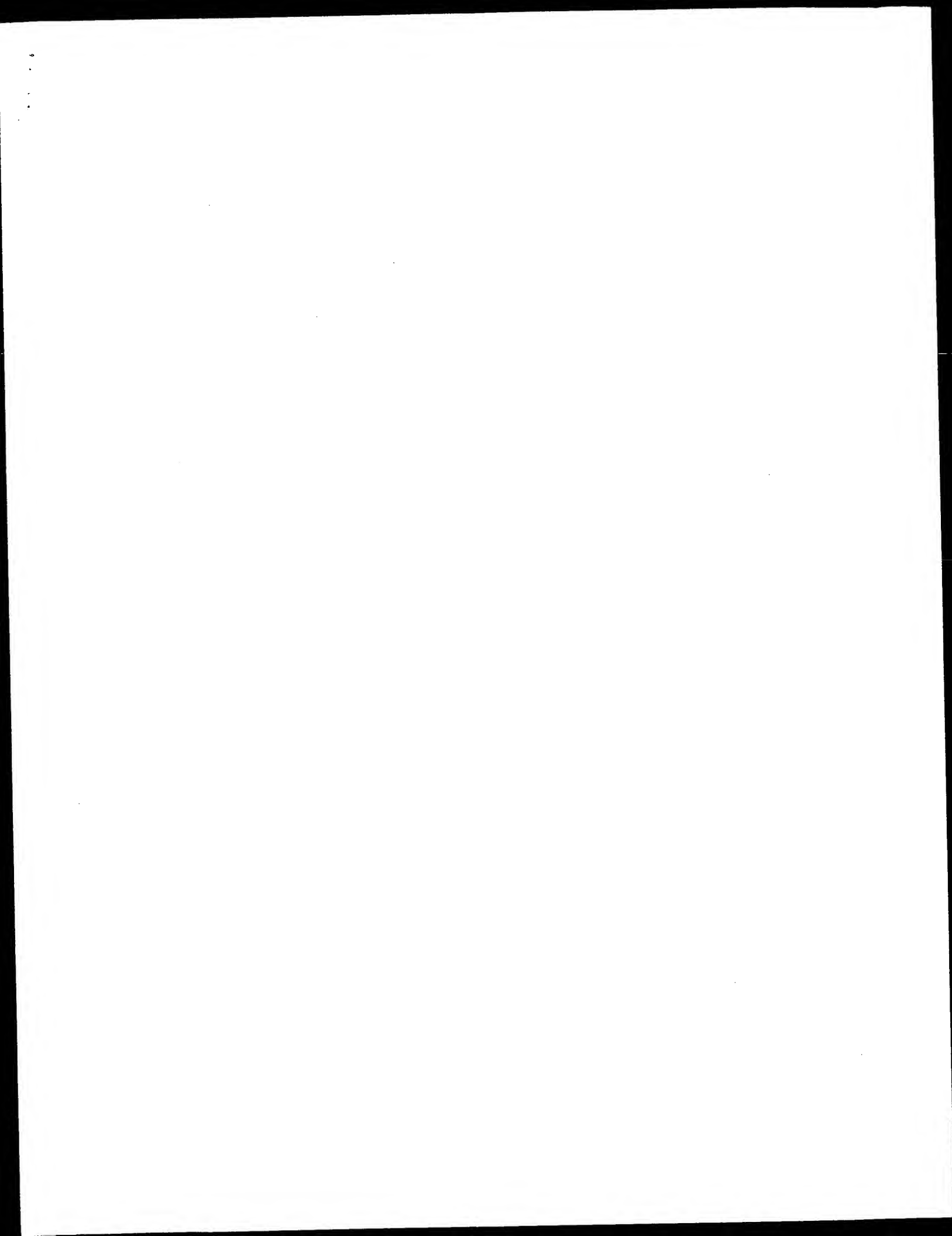
SO Sequence 831 BP; 125 A; 99 C; 281 G; 326 T; 0 other;

Query Match 78.2%; Score 17.2; DB 24; Length 831;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACACGAA 22
Db 771 AAAAAAAATCTAACAACGAA 750

Search completed: March 17, 2003, 10:50:47
Job time: 128.253 secs



GenCore version 5.1.4-P5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 777.688 Seconds

(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-4

Perfect score: 22

Sequence: 1 aagaaaaatcagaacagca 22

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST.*
1: em_estb1:*
2: em_estb2:*
3: em_estb3:*
4: em_estb4:*
5: em_estb5:*
6: em_estb6:*
7: em_estb7:*
8: em_estb8:*
9: em_estb9:*
10: em_estb10:*
11: em_estb11:*
12: em_estb12:*
13: em_estb13:*
14: em_estb14:*
15: em_estb15:*
16: em_estb16:*
17: em_estb17:*
18: em_estb18:*
19: em_estb19:*
20: em_estb20:*
21: em_estb21:*
22: em_estb22:*
23: em_estb23:*
24: em_estb24:*
25: em_estb25:*
26: em_estb26:*
27: em_estb27:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.4	92.7	342	10	AM740409
2	20.4	92.7	342	12	BS360474
3	20.4	92.7	360	10	AM740410
4	18.8	85.5	224	10	BB016617
5	18.8	85.5	424	17	BB670801
6	18.8	85.5	449	17	AA009207

C	7	18.8	85.5	476	10	BB766795
C	8	18.8	85.5	505	12	BF004802
C	9	18.8	85.5	528	19	AA447445
C	10	18.8	85.5	544	17	AA014714
C	11	18.8	85.5	549	17	BB282218
C	12	18.8	85.5	567	17	AA0534621
C	13	18.8	85.5	568	17	AA0534621
C	14	18.8	85.5	597	17	AA0534617
C	15	18.8	85.5	710	17	BB552202
C	16	18.8	85.5	738	17	AA572374
C	17	18.8	85.5	767	17	CNS070709
C	18	18.8	85.5	1059	17	CNS06808
C	19	18.8	85.5	1330	12	BF164624
C	20	18.4	83.6	479	17	CNS01001
C	21	18.8	81.8	507	10	AM828026
C	22	18.8	81.8	582	10	AM871607
C	23	18.8	81.8	597	10	AM871607
C	24	18.8	81.8	624	10	AM870709
C	25	18.8	81.8	651	10	AM735500
C	26	17.8	80.9	229	10	AV380205
C	27	17.8	80.9	280	10	BB381647
C	28	17.8	80.9	283	10	BB106733
C	29	17.8	80.9	318	9	AL837465
C	30	17.8	80.9	318	13	BI034283
C	31	17.8	80.9	501	10	AV873095
C	32	17.8	80.9	510	17	AZ035704
C	33	17.8	80.9	527	10	AV873962
C	34	17.8	80.9	533	14	BO234625
C	35	17.8	80.9	539	17	BA427923
C	36	17.8	80.9	558	10	AM034365
C	37	17.8	80.9	632	17	A2417128
C	38	17.8	80.9	638	10	AV872452
C	39	17.8	80.9	656	17	BB059762
C	40	17.8	80.9	666	10	BB641601
C	41	17.8	80.9	666	17	BB717507
C	42	17.8	80.9	670	17	DR10N245
C	43	17.8	80.9	703	10	AV868791
C	44	17.8	80.9	775	17	BB705808
C	45	17.8	80.9	781	17	BB059760

ALIGNMENTS

RESULT 1
LOCUS AM740409 342 bp mRNA linear EST 27-APR-2000
DEFINITION BR10552 Blomphalaria glabrata (BS-90)-unexposed lambda zap library
ACCESSION Blomphalaria glabrata cDNA clone RBG1155TR, mRNA sequence.
VERSION AM740409.1 GI:7651688
KEYWORDS EST.

ORGANISM

Blomphalaria planorb.
Blomphalaria glabrata
Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora
Planorbidae; Blomphalaria.

REFERENCE

1 (bases 1 to 342)
Raghavan, N., Miller, A., Gardner, M., Kerlavage, A.R., Fitzgerald, P.C.,
Lewis, F.A., and Knight, M.
Genes expressed by the hemocytes of Blomphalaria glabrata before
and after exposure to miracidia

JOURNAL

Unpublished (2000)
Contact: Raghavan N
Biomedical Research Institute
1211 Parklawn Dr., Rockville, MD 20852, USA
Tel: 301-881-3300 ext.128
Fax: 301-770-4756
Email: nkr@helix.nih.gov, snallstrule@aol.com.

FEATURES

source
1..342
Location/Qualifiers
/organism="Blomphalaria glabrata"
/strain="BS-90"
/db_xref="taxon:6526"

```

/clone="RBGIH55TR"
/clone.lib="Biomphalaria glabrata (BS-90) -unexposed Lambda
zap library"
/sex="hermaphrodite"
/cell_type="Hemocyte"
/lab_host="Laboratory host"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Total RNA was isolated from the hemocytes of
unexposed Biomphalaria glabrata (BS-90) snails and first
strand cDNA synthesized using an oligo-dT primer-linker
(XhoI). Second strand synthesis was followed by the
ligation of EcoRI adaptors. Following digestion with XhoI,
the completed, directional cDNA was cloned into Uni-ZAP
XR phagemid vector by Stratagene."
BASE COUNT      132 a      68 c      62 g      80 t
ORIGIN
Query Match      92.7%; Score 20.4; DB 10; Length 342;
Best Local Similarity 95.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 AAGAAAAATCTAGACAGCAA 22
|||||
Db      174 AAGAAAAATCTAGACAGCAA 195

RESULT 2
BG360474      342 bp      mRNA      linear      EST 07-MAR-2001
LOCUS      BR110646 Biomphalaria glabrata (BS-90) -unexposed Lambda zap library
DEFINITION      Biomphalaria glabrata cDNA clone RBGIH85TR, mRNA sequence.
ACCESSION      BG360474
VERSION      BG360474.1 GI:13243488
KEYWORDS      EST.
SOURCE      bloodfluke planorb.
ORGANISM      Biomphalaria glabrata
Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora
; Planorbidae; Biomphalaria.
REFERENCE      1 (bases 1 to 342)
AUTHORS      Raghavan,N., Miller,A., Gardner,M., Kerlavage,A.R., Fitzgerald,P.C.,
Lewis,F.A. and Knight,M.
Genes expressed by the hemocytes of Biomphalaria glabrata before
and after exposure to miracidia
Unpublished (2000)
JOURNAL      Biomedical Research Institute
COMMENT      12111 Parklawn Dr., Rockville, MD 20852, USA
Tel: 301-881-3300 ext.128
Fax: 301-770-4756
Email: nkr@helix.nih.gov, snailstrule@aol.com.
FEATURES
source      1. .342
Location/Qualifiers
/organism="Biomphalaria glabrata"
/strain="BS-90"
/db_xref="taxon:6526"
/clone="RBGIH85TR"
/clone.lib="Biomphalaria glabrata (BS-90) -unexposed Lambda
zap library"
/sex="hermaphrodite"
/cell_type="Hemocyte"
/lab_host="Laboratory host"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Total RNA was isolated from the hemocytes of
unexposed Biomphalaria glabrata (BS-90) snails and first
strand cDNA synthesized using an oligo-dT primer-linker
(XhoI). Second strand synthesis was followed by the
ligation of EcoRI adaptors. Following digestion with XhoI,
the completed, directional cDNA was cloned into Uni-ZAP
XR phagemid vector by Stratagene."
BASE COUNT      133 a      68 c      63 g      78 t
ORIGIN
Query Match      92.7%; Score 20.4; DB 12; Length 342;

```

```

Best Local Similarity 95.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 AAGAAAAATCTAGACAGCAA 22
|||||
Db      174 AAGAAAAATCTAGACAGCAA 195

RESULT 3
AM740410      360 bp      mRNA      linear      EST 27-APR-2000
LOCUS      BR110553 Biomphalaria glabrata (BS-90) -unexposed Lambda zap library
DEFINITION      Biomphalaria glabrata cDNA clone RBGIH56TR, mRNA sequence.
ACCESSION      AM740410
VERSION      AM740410.1 GI:7651689
KEYWORDS      EST.
SOURCE      bloodfluke planorb.
ORGANISM      Biomphalaria glabrata
Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora
; Planorbidae; Biomphalaria.
REFERENCE      1 (bases 1 to 360)
AUTHORS      Raghavan,N., Miller,A., Gardner,M., Kerlavage,A.R., Fitzgerald,P.C.,
Lewis,F.A. and Knight,M.
Genes expressed by the hemocytes of Biomphalaria glabrata before
and after exposure to miracidia
Unpublished (2000)
JOURNAL      Biomedical Research Institute
COMMENT      12111 Parklawn Dr., Rockville, MD 20852, USA
Tel: 301-881-3300 ext.128
Fax: 301-770-4756
Email: nkr@helix.nih.gov, snailstrule@aol.com.
FEATURES
source      1. .360
Location/Qualifiers
/organism="Biomphalaria glabrata"
/strain="BS-90"
/db_xref="taxon:6526"
/clone="RBGIH56TR"
/clone.lib="Biomphalaria glabrata (BS-90) -unexposed Lambda
zap library"
/sex="hermaphrodite"
/cell_type="Hemocyte"
/lab_host="Laboratory host"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Total RNA was isolated from the hemocytes of
unexposed Biomphalaria glabrata (BS-90) snails and first
strand cDNA synthesized using an oligo-dT primer-linker
(XhoI). Second strand synthesis was followed by the
ligation of EcoRI adaptors. Following digestion with XhoI,
the completed, directional cDNA was cloned into Uni-ZAP
XR phagemid vector by Stratagene."
BASE COUNT      137 a      72 c      62 g      86 t      3 others
ORIGIN
Query Match      92.7%; Score 20.4; DB 10; Length 360;
Best Local Similarity 95.5%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 AAGAAAAATCTAGACAGCAA 22
|||||
Db      174 AAGAAAAATCTAGACAGCAA 195

RESULT 4
BB016617      224 bp      mRNA      linear      EST 22-JUN-2000
LOCUS      BB016617 RIKEN full-length enriched, adult male testis (DH10B) Mus
DEFINITION      musculus cDNA clone 4930563D02 3', mRNA sequence.
ACCESSION      BB016617
VERSION      BB016617.1 GI:8187765
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus

```

TITLE	JOURNAL	COMMENT
RIKEN MOUSE ESTS (Konno, H., et al.)	Unpublished (2000)	Contact: Yoshitohide Hayashizaki

FEATURES
SOURCE

Query Match

85.5%; Score 18.8; DB 10; Length 224;

99 a 28 c 39 g 58 t

```
OY      1 AAGAAAAAATCTAGACAAGCAA 22  
         ||||| | | | | |  
Db     160 AAGAAAAAATTCAGACAAGCAA 181
```

Accession	Length	Source	Definition
U00000001	424 bp	DNA	linear
BOMLRK25TF	BO_2_3_KB	Brassica oleracea genomic clone	BOMLRK25, DNA sequence.

SOURCE	ORGANISM
Brassica oleracea.	Brassica oleracea

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 424)
Town, C.D., Van Aken, S., Uterback, T., and Fraser, C.M.
Whole genome shotgun sequencing of *Brassica oleracea*
Unpublished (2001)
Contact: Chris Town

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtowne@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES	Location/Qualifiers
source	1. .424

```

/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOMLR25"
/clone_1lb="BO_2_3_KB"
/notes="Vector: pHS01; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHS01 using BstXI linkers"
BASE COUNT
165 a      86 c      66 g      107 t
BRIGIN

```

Query Match	85.5%	Score 18.8;	DB 17;	Length 424;
Best Local Similarity	90.9%	Pred. No. 3.8e+03;		
Matches 20; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY 1 AAGGAAAAATCTAGACAAACCA 22
||| ||||| ||||| |||||
Db 127 AAGTAAAAATCTAGATAAGCAA 148

RESULT 6	449 bp	DNA	linear	GSS 25-FEB-2000
AZ009207/c				
LOCUS				
DEFINITION				
ACCESSION				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				
101				
102				
103				
104				
105				
106				
107				
108				
109				
110				
111				

```

VERSION      AZ0092207.1  GI:7084591
KEYWORDS
SOURCE
GSS.         house mouse.

```

REFERENCE
1 (bases 1 to 449)
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)

COMMENT

Other GSSS: RPCI-23-365C18.TU
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0200

Email: szhaoc@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@edjlong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.html) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 365 row: C column: 18
 Seq primer: 17
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..449
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-365C18"
 /clone.lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; EcoRI: Site: 2; EcoRI: Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
 79 c 80 g 193 t

BASE COUNT

97 a 79 c 80 g 193 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 17; Length 449;
 Best Local Similarity 90.9%; Pred. No. 3.8e+03;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACACGCA 22

Db 423 AATAAAAATCTAGACACGCA 402

RESULT 7
 BB768795/c 476 bp mRNA linear EST 17-OCT-2001
 LOCUS BB768795 RIKEN full-length enriched, B16 F10Y cells Mus musculus
 DEFINITION cDNA clone G370089D24 3', mRNA sequence.
 ACCESSION BB768795
 VERSION BB768795.1 GI:16211337
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 476)
 Akimura,T., Arakawa,T., Carinici,P., Furuno,M., Hanaagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirazane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shitaki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE

JOURNAL

Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)

FEATURES

source

Location/Qualifiers
 1..476
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370089D24"
 /clone.lib="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 /note="pooled tissues; (tissue_type=cerebellum, dev_stage=16 days neonate, sex-mixed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex-mixed), (tissue_type=hippocampus, dev_stage=adult, sex-male), (tissue_type=whole body, dev_stage=9 days embryo, sex-mixed), (tissue_type=lung, dev_stage=13 days embryo, sex-mixed)"
 e mouse tissues.
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

BASE COUNT

128 a 105 c 111 g 132 t

ORIGIN

Query Match 90.5%; Score 18.8; DB 10; Length 476;
 Best Local Similarity 95.9%; Pred. No. 3.8e+03;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTAGACACGCA 22

Db 303 AAAAAAAACCTAGACACGCA 282

RESULT 8
 BF004802/c 505 bp mRNA linear EST 06-OCT-2000
 LOCUS BF004802 KVL Medicago truncatula cDNA clone pVL1-18B14, mRNA
 DEFINITION sequence.
 ACCESSION BF004802
 VERSION BF004802.1 GI:10705077
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
 1 (bases 1 to 505)
 Vandenbosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,L.E. and Fraser,C.M.
 Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,L.E. and Fraser,C.M.
 ESTs from roots of Medicago truncatula 24 hours after inoculation with Sinorhizobium meliloti
 Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenbosch.umn.edu
Texas A&M University name: T268612e TIGR sequence name: MTIBJ077K
More information is available at: <http://chysie.tamu.edu/medicago>
Seq primer: Skmod (CTA GAA CTA gta gta CC).
Location/Qualifiers

FEATURES

source

1. 505
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="PKV1-18B14"
/clone_1lb="KVI"
/lissue_type="Seedling roots"
/dev_stage="24 hours post-inoculation with Sinorhizobium
meliloti"
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the unizap XR vector from
stratagene and packaged using gigapack iii gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XLOLR cells."
1 others

BASE COUNT
ORIGIN

144 a 93 c 103 g 164 t

Query Match

Best Local Similarity 90.9%; Score 18.8; DB 12; Length 505;
Pred. No. 3.7e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTGACACGAA 22
||||| ||||| ||||| |||||

Db 109 AAGAAAAATCTGACACGAA 88

RESULT 9
AA447445

LOCUS zw93g12.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
DEFINITION IMAGE:784562 5', mRNA sequence.

AA447445
AA447445 1 GI:2161115

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 528)
Hillier, L., Allen, M., Bowles, L., Dubouche, T., Giesel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wyllie,
T., Waterston, R. and Wilson, R.
Washu-Merck EST Project 1997
Unpublished (1997)

TITLE
JOURNAL
COMMENT

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28m3 rev2 ET from Amersham
High quality sequence stop: 475.
Location/Qualifiers

FEATURES

source

1. 528
/organism="Homo sapiens"
/db_xref="GDB:598220"

/db_xref="taxon:9606"
/clone="IMAGE:784582"
/clone_1lb="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
GTGACCAATCTGAGTGGAGCGCCGCTTAATTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Benito Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

155 a 113 c 91 g 169 t

Query Match

Best Local Similarity 90.9%; Score 18.8; DB 9; Length 528;
Pred. No. 3.7e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGAAAAATCTGACACGAA 22
||||| ||||| ||||| |||||

Db 24 AAGAAAAATCTGACACGAA 45

RESULT 10
A0414714

LOCUS RPCI-11-171019.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-171019
DEFINITION 'DNA sequence.

ACCESSION A0414714
VERSION A0414714.1 GI:4473683

KEYWORDS
SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 544)
Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter,
J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
Unpublished (1997)

Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208

Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: SP6
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1. 544
/organism="Homo sapiens"
/db_xref="GDB:7565634"
/db_xref="taxon:9606"
/clone="RPCI-11-171019"
/clone_1lb="RPCI-11"
/sex="Male"

BASE COUNT

ORIGIN

189 a 97 c 100 g 158 t

/cell_type="Lymphocytes"
/note="Vector: pACE6.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"

BASE COUNT
ORIGIN

189 a 97 c 100 g 158 t

BASE COUNT
ORIGIN

189 a 97 c 100 g 158 t

Query Match	Similarity	85.5%	Score	18.8	DB	17	Length	544	
Best Local	Similarity	90.9%	Pred. No.	3.7e+03					
Matches	20	Conservative	0	Mismatches	2	Indels	0	Gaps	0
QY	1	AAAGAAAAATCTGACACAGCAA	22						
Db	488	AAAGAAAAAGCTAGACAAAGAA	509						
RESULT 11									
LOCUS	BH282218/c								
DEFINITION	BH282218		549 bp	DNA	linear	GSS	30-NOV-2001		
ACCESSION	CH230-39A11								
VERSION	CH230-39A11								
KEYWORDS	BH282218								
SOURCE	GSS.								
ORGANISM	GI:17194620								
	Norway rat.								
	Rattus norvegicus								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;								
	Ratus.								
REFERENCE	1 (bases 1 to 549)								
AUTHORS	Zhao,S., Shelly,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn								
	,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de								
	Jong,P., and Fraser,C.M.								
TITLE	Rat BAC End Sequences from Library CHORI-230 EcoRI segment								
JOURNAL	Unpublished (1999)								
COMMENT	Other_GSS: CH230-39A11.TJ								
	Contact: Shaying Zhao								
	Department of Eukaryotic Genomics								
	The Institute for Genomic Research								
	9712 Medical Center Dr., Rockville, MD 20850, USA								
	Tel: 301 838 0200								
	Fax: 301 838 0208								
	Email: szhao@tigr.org								
	Clones are derived from the rat BAC library CHORI-230								
	(http://www.chori.org/bacpac/fat230.htm). For BAC library								
	availability, please contact Pieter de Jong (pjejong@mail.cho.org)								
	Clones may be purchased from BACPAC Resources								
	(http://www.chori.org/bacpac/or_ering_information.htm). BAC end								
	page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html								
	Plate: 39 row: A column: 11								
	Seq primer: T7								
	Class: BAC ends.								
FEATURES									
SOURCE	Location/Qualifiers								
	1. 549								
	/organism="Rattus norvegicus"								
	/strain="BN/SSNHsd/MCw"								
	/db_xref="taxon:10116"								
	/clone="CH230-39A11"								
	/clone_lib="CHORI-230 Segment 1"								
	/sex="Female"								
	/cell_type="Brain"								
	/note="Vector: pPRBAC2.1; Site_1: EcoRI; site_2: EcoRI;								
	CHORI-230 Rat (BN/SSNHsd/MCw) BAC library produced by								
	Pieter de Jong"								
BASE COUNT	118 a 106 c 125 g 200 t								
ORIGIN									
Query Match		85.5%	Score	18.8	DB	17	Length	549	
Best Local	Similarity	90.9%	Pred. No.	3.7e+03					
Matches	20	Conservative	0	Mismatches	2	Indels	0	Gaps	0
QY	1	AAAGAAAAATCTAGACAGCAA	22						
Db	43	AAAGAAAAAGCTAGACAAACAA	22						
RESULT 12									
LOCUS	AO534621								
COMMENT									
	567 bp								
	DNA								
	linear								
	GSS								
	18-MAY-1999								

DEFINITION	RPCI-11-353P23.TV RPCI-11 Homo sapiens genomic clone RPCI-11-353P23 DNA sequence.
ACCESSION	A0534621
VERSION	A0534621.1 GI:4846311
KEYWORDS	GSS.
SOURCE ORGANISM	Homo sapiens human. <i>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.</i>
REFERENCE AUTHORS	1 (bases 1 to 567) Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter, J.C.
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL COMMENT	Unpublished (1997)
Other_GSSTS:	RPCI-11-353P23.TV
Contact:	Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel.: 301 838 0200 Fax: 301 838 0208 Email: hbeetlgr.org
Clares are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@rsngen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html . Seq primer: 17	
Class:	BAC ends.
FEATURES source	Location/Qualifiers
	1..567
/organism=	"Homo sapiens"
/db_xref=GDB:	7635550"
/db_xref=taxon:	9606"
/clone=RPCI-	11-353P23"
/clone.lib=RPCI-	11"
/sex=Male"	
/cell_type=Lymphocytes"	
/note=vector:	PBACC3.6; Site_1: EcoRI; site_2: EcoRI;
RPCII Human Male BAC Library"	
BASE COUNT	191 a 99 c 111 g 165 t . 1 others
ORIGIN	
Query Match	85.5%; Score 18.8; DB 17; Length 567;
Best Local Similarity	90.9%; Pred. No. 3.7e+03;
Matches	20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY	1 AAGAAAAATCTGACACAGCA 22
Db	476 AAGAAAAGAAGTGAACAAGAA 497
RESULT 13	
LOCUS	AZ506877 568 bp DNA linear GSS 05-OCT-2000
DEFINITION	IM0348C14F Mouse 10kb plasmid UGCJM library Mus musculus genomic clone UUGCJM0348C14 F, DNA sequence.
ACCESSION	AZ506877
VERSION	AZ506877.1 GI:10688193
KEYWORDS	GSS.
SOURCE ORGANISM	house mouse. <i>Mus musculus</i> <i>Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.</i>
REFERENCE	1 (bases 1 to 568) Barber,M., Beacroft,T., Duval,B., Hamill,C., Dunn,D., Aoysgi,A., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly ,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.
AUTHORS	Mouse whole genome scaffolding with paired end reads from 10Kb plamid inserts
TITLE	

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0348 row: C column: 14
Seq primer: CGTTGTAACGACGCCCACT
Class: plasmid ends
High quality sequence step: 568.
Location/Qualifiers
1..568

FEATURES

Location/Qualifiers
1. .568

BASE COUNT	192 a	101 c	123 g	150 t	2 others
ORIGIN					

Query Match	85.5%;	Score 18.8;	DB 17;	Length 568;
Best Local Similarity	90.9%;	Pred. No. 3.7e+03;		
Matches 20; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

Oy	1	AAGAA	AAATCTAGACA	GCAA	22
Db	191	AAGAA	AAGATCTAGACA	GAA	212

RESULT 14
A0534617

Accession	U05340.1	597 bp	DNA	linear	GSS 18-MAY-1999
Definition	HPC1-11 Homo sapiens genomic clone HPC1-11-353p21, DNA sequence.				

ACCESSION	A0534617
VERSION	A0534617.1
KEYWORDS	GI:4846307
SOURCE	GSS.
ORGANISM	human.
	Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 597)
Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.

JOURNAL COMMENT	map Building unpublished (1997) Other_GSSS: RPCI-11-353P21.TJ
-----------------	--

Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0200

Email: hbeef@igf.org
 Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pierer de Jong (pierre@dejong.med.buffalo.edu). Clones may be purchased from BACBPC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (inforesgen.com). BAC end search page: http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: 17
 Class: BAC ends.

FEATURES

Source

```

/organism="Homo sapiens"
/db_xref="GDB:7635548"
/cb_xref="taxon:9606"
/clone="RPC1-11-353p21"
/clone_1b="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI
RPC11 Human Male BAC Library"

```

Query Match	85.5%	Score 18.8	DB 17	Length 597
Best Local Similarity	90.9%	Pred. No. 3.6e+03		
Matches	20	Conservative	0	Mismatches 2
				Indels 0
				Gaps 0

QY	1	AAGAAAAAATCTAGACAAGCAA	22
Db	496	AAGAAAAAAGCTAGACAAGGAA	517

RESULT 15
BH552202/

LOCUS	BH552202	710 bp	DNA	linear	GSS 14-DEC-2001
DEFINITION	BOHQN17TF BOHQ Brassica oleracea genomic clone BOHQN17, DNA sequence.				

VERSION BH552202.1 GI:17803982
KEYWORDS GSS.
SOURCE Brassica oleracea.

REFERENCE
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 / bases 1 to 710

TITLE Whole genome shotgun sequencing of *Brassica oleracea*
 JOURNAL Unpublished (2001)
 COMMENT Other_GSSs: BOH0N17TR

TIGR
9712 Medical Center Dr
Tel: 301-838-3523
Fax: 301-838-0000

DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES

Source

```

/organism="Brassica oleracea"
/strain="T01000DH1"
/db_xref="taxon:3712"
/clone="BOH0N17"
/clone_1lb="BOH0"
/note="Vector: pMOSt. Site_1: BstXI; 2-3 kb sheared

```

BASE COUNT genomic DNA inserted into PHOS1 using BstXI linkers"
ORIGIN 209 a 146 c 149 g 206 t

Query Match 85.5%; Score 18.8; DB 17; Length 710;
Best Local Similarity 90.9%; Pred. No. 3.5e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGAAAAATCTAGACAAAGCAA 22
 |||| ||||||||| |||||
Db 154 AAGATTAATCTAGATAAGCAA 133

Search completed: March 17, 2003, 13:09:20
Job time : 782.688 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 230.108 Seconds

(without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-5

Perfect score: 25

Sequence: 1 gcttcttgcgtcagagcttcca 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_da:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_cm:*
- 5: gb_ov:*
- 6: gb_pal:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_pro:*
- 11: gb_scs:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_cm:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_scs:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hlg_hum:*
- 31: em_hlg_inv:*
- 32: em_hlg_other:*
- 33: em_hlg_mus:*
- 34: em_hlg_pln:*
- 35: em_hlg_rtd:*
- 36: em_hlg_mam:*
- 37: em_hlg_vrt:*
- 38: em_sy:*
- 39: em_hlgo_hum:*
- 40: em_hlgo_mus:*
- 41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.4	93.6	269	10 MUSOPS4	M36698 Mouse opsin
2	23.4	93.6	2610	10 BC031766	BC031766 Mus muscu
3	23.4	93.6	3249	10 BC013125	BC013125 Mus muscu
4	23.4	93.6	9483	10 MUSOPS	M55171 Mouse opsin
5	21.8	87.2	175	5 AF249155	AF249155 Indirana
6	21.8	87.2	1062	5 TN1293018	AJ299018 Tetradon
7	20.4	81.6	178639	2 AC095437	AC095437 Rattus no
8	20.2	80.8	169	5 AF249131	AF249131 Microhya
9	20.2	80.8	175	5 AF249129	AF249129 Bufo meli
10	20.2	80.8	175	5 AF249133	AF249133 Mantella
11	20.2	80.8	175	5 AF249134	AF249136 Mantidact
12	20.2	80.8	175	5 AF249136	AF249136 Boophis x
13	20.2	80.8	175	5 AF249137	AF249137 Boophis t
14	20.2	80.8	175	5 AF249138	AF249138 Laliostom
15	20.2	80.8	175	5 AF249139	AF249139 Refervary
16	20.2	80.8	175	5 AF249140	AF249140 Refervary
17	20.2	80.8	175	5 AF249141	AF249141 Hoplobatr
18	20.2	80.8	175	5 AF249142	AF249142 Sphaerotr
19	20.2	80.8	175	5 AF249143	AF249143 Euphyctyl
20	20.2	80.8	175	5 AF249144	AF249144 Nannophry
21	20.2	80.8	175	5 AF249146	AF249146 Nyctibatr
22	20.2	80.8	175	5 AF249147	AF249147 Limnodyn
23	20.2	80.8	175	5 AF249148	AF249148 Limnodyn
24	20.2	80.8	175	5 AF249149	AF249149 Rana curt
25	20.2	80.8	175	5 AF249150	AF249150 Rana temp
26	20.2	80.8	175	5 AF249151	AF249151 Rana temp
27	20.2	80.8	175	5 AF249152	AF249152 Microxalu
28	20.2	80.8	175	5 AF249154	AF249154 Indirana
29	20.2	80.8	175	5 AF249156	AF249156 Polypedat
30	20.2	80.8	175	5 AF249157	AF249157 Rhacophor
31	20.2	80.8	175	5 AF249158	AF249158 Philautus
32	20.2	80.8	405	5 AF221974	AF221974 Bufo pusi
33	20.2	80.8	416	5 AF221979	AF221979 Bufo pant
34	20.2	80.8	825	5 AF137213	AF137213 Ostracion
35	20.2	80.8	924	5 AB084931	AB084931 Dimidioc
36	20.2	80.8	924	5 AB084938	AB084938 Oreochrom
37	20.2	80.8	924	5 AB084940	AB084940 Sarothero
38	20.2	80.8	924	5 AB084941	AB084941 Spatiodus
39	20.2	80.8	924	5 AB084944	AB084944 Tilapia r
40	20.2	80.8	924	5 AB084947	AB084947 Xenotilap
41	20.2	80.8	1053	5 AF021242	AF021242 Melopsitt
42	20.2	80.8	1059	5 AMU12328	U12328 Aslyvmax me
43	20.2	80.8	1399	5 RTU59920	RTU59920 Rana temp
44	20.2	80.8	1423	5 MSU118666	Y18666 Mullus surm
45	20.2	80.8	1469	5 MVU57539	Y57539 Myrtilpistis

ALIGNMENTS

RESULT 1	MUSOPS4	269 bp	DNA	Linear	ROD 08-MAY-1993
LOCUS	MUSOPS4				
DEFINITION	Mouse opsin gene, exon 4.				
ACCESSION	M36698 X69175				
VERSION	M36698.1 GI:200149				
KEYWORDS	opsin.				
SEGMENT	4 of 5				
SOURCE	Mus musculus (strain C56BL/6J) eye DNA.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Baehr, W., Falk, J.D., Bugra, K., Triantafyllou, J.T. and McGinnis, J.F.				
TITLE	Isolation and analysis of the mouse opsin gene				

JOURNAL FEBS Lett. 238 (2), 253-256 (1988)
MEDLINE 89005694
PUBMED 2844600

FEATURES
SOURCE Location/Qualifiers

1..269
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/tissue_type="eye"
order(M36897.1:170, >189,<1, .14)
/gene="Opsin"
/note="117 bp gap"
/number=3
15..254
/gene="opsin"
/number=4

BASE COUNT 57 a 87 c 58 g 67 t
ORIGIN About 117 bp after segment 3.

Query Match 93.6%; Score 23.4; DB 10; Length 269;
Best Local Similarity 96.0%; Pred. No. 1.4;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCTTCTTTGCTGAGAGCTTCCA 25
|||||
DB 192 GCTTCTTTGCTGAGAGCTTCCA 216

RESULT 2
BC031766 2610 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, clone MGC:25387 IMAGE:4527040, mRNA, complete cds.
DEFINITION BC031766
ACCESSION BC031766.1 GI:21594394
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnarste, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAC Plate: 31 Row: n Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction.

FEATURES
SOURCE Location/Qualifiers

1..2610
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:25387 IMAGE:4527040"
/tissue_type="eye, retina, mouse strain C57BL/6"

CDS
/clone_11b="NIH_MGC_94"
/lab_host="DH10B"
/note="vector: PCMV-SPORT6"
80..1126
/codon_start=1
/product="unknown (protein for MGC:25387)"
/protein_id="AAH31766.1"
/db_xref="GI:21594395"
/db_xref="locustid:212541"

BASE COUNT 604 a 727 c 656 g 623 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 10; Length 2610;
Best Local Similarity 96.0%; Pred. No. 1.5;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCTTCTTTGCTGAGAGCTTCCA 25
|||||
DB 953 GCTTCTTTGCTGAGAGCTTCCA 977

RESULT 3
BC013125 3249 bp mRNA linear ROD 07-AUG-2002
LOCUS Mus musculus, similar to rhodopsin (opsin 2, rod pigment) (retinitis pigmentosa 4, autosomal dominant), clone MGC:21565
DEFINITION BC013125
ACCESSION BC013125.1 GI:15341884
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnarste, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
Series: IRAC Plate: 28 Row: e Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES
SOURCE Location/Qualifiers

1..3249
/organism="Mus musculus"
/db_xref="taxon:10090"

REFERENCE 1 (bases 1 to 175)
 AUTHORS Bossuyt, F., and Milinkovitch, M.C.
 TITLE Convergent Adaptive Radiations in Madagascar and Asian Rain Forests
 JOURNAL Reveal Co-variation between Larval and Adult Traits
 REFERENCE 2 (bases 1 to 175)
 AUTHORS Bossuyt, F., and Milinkovitch, M.C.
 TITLE Direct Submission
 JOURNAL Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of Molecular Biology and Medicine, rue Jeener and Brachet 12, Gosselies B-6041, Belgium

FEATURES
 source Location/Qualifiers
 1..175
 /organism="Indirana sp. 2"
 /db_xref="taxon:147866"
 mRNA <1..>175
 /product="rhodopsin"
 CDS <1..>175
 /codon_start=2
 /product="rhodopsin"
 /protein_id="AG49798.1"
 /db_xref="GI:12247240"
 /translation="ARKEVTRMVMVWVFLICWVRYAVAVYFIHQSEFGPIFMT
 VPAFFAKSSSIYNP"

BASE COUNT 34 a 54 c 37 g 49 t 1 others
 ORIGIN
 exon
 Query Match 87.2%; Score 21.8; DB 5; Length 175;
 Best Local Similarity 92.0%; Pred. No. 8.9;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCCTTGCTGAGACCTCTTCCA 25
 Db 140 GCCTTCCTTGCTGAGACCTCTTCCA 164

RESULT 6
 LOCUS TN1293018 1062 bp DNA linear VRT 23-AUG-2000
 DEFINITION Tetraodon nigroviridis rhod gene for rhodopsin.
 ACCESSION AJ293018
 VERSION AJ293018.1 GI:9909983
 KEYWORDS RHOD gene; rhodopsin.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 1062)
 Fischer, C.
 REFERENCE 1
 TITLE Tetraodon nigroviridis gene for rhodopsin
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1062)
 AUTHORS Fischer, C.
 TITLE Direct Submission
 JOURNAL Submitted (22-AUG-2000) Fischer C., Centre National de Sequençage, Genoscope, 2 rue Gaston Crémieux, CP5706, F-91057 Evry Cedex, FRANCE

FEATURES
 source Location/Qualifiers
 1..1062
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 gene 1..1062
 /gene="rhod"
 CDS 1..1062
 /gene="rhod"
 /codon_start=1
 /product="rhodopsin"
 /protein_id="CAC04526.1"
 /db_xref="GI:9909984"

REFERENCE 1 (bases 1 to 178639)
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaralunga, H.C., Are, J.R., Ayele, M., Banks, T., Barbarella, J., Benon, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Bunyah, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.P., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loutseged, H., Lozato, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newson, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokem, S., Oghu, M., Okwunodu, G., Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, N., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Slisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabor, P., Tamerisa, K., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Wortley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

BASE COUNT 198 a 329 c 259 g 276 t
 ORIGIN
 Query Match 87.2%; Score 21.8; DB 5; Length 1062;
 Best Local Similarity 92.0%; Pred. No. 8.9;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTTCCTTGCTGAGACCTCTTCCA 25
 Db 874 GCCTTCCTTGCTGAGACCTCTTCCA 898

RESULT 7
 LOCUS AC095437/c 178639 bp DNA linear HTG 10-JUL-2002
 DEFINITION Rattus norvegicus c1one CH230-4H9, *** SEQUENCING IN PROGRESS ***
 ACCESSION AC095437
 VERSION AC095437.3 GI:21716847
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

1 (bases 1 to 178639)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaralunga, H.C., Are, J.R., Ayele, M., Banks, T., Barbarella, J., Benon, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Bunyah, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.P., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loutseged, H., Lozato, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newson, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokem, S., Oghu, M., Okwunodu, G., Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, N., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Slisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swalek, A., Tabor, P., Tamerisa, K., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Wortley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 178639)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 178639)
Worley, K.C.
Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 9, 2002 this sequence version replaced gi:17941914.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: CGHA
Center clone name: CH230-4H9

----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 128571 bases at least Q40
Consensus quality: 133786 bases at least Q30
Consensus quality: 137554 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1037 1036: contig of 1036 bp in length
1137 1136: gap of unknown length
2378 2377: contig of 1241 bp in length
2478 2477: gap of unknown length
3635 3634: contig of 1157 bp in length
3735 3734: gap of unknown length
5028 5027: contig of 1293 bp in length
5128 5127: gap of unknown length
6328 6327: contig of 1200 bp in length
6428 6427: gap of unknown length
8035 8034: contig of 1607 bp in length
8135 8134: gap of unknown length
9425 9424: contig of 1290 bp in length
9525 9524: gap of unknown length
10939 10938: contig of 1414 bp in length
12078 12077: gap of unknown length
12178 12177: contig of 1039 bp in length
13191 13190: gap of unknown length
13291 13290: gap of unknown length
14395 14394: contig of 1104 bp in length
14495 14494: gap of unknown length
15667 15666: contig of 1172 bp in length
15767 15766: gap of unknown length
17029 17028: contig of 1263 bp in length
17130 17129: gap of unknown length
18245 18244: contig of 1115 bp in length
18345 18344: gap of unknown length
19988 19987: contig of 1643 bp in length
20088 20087: gap of unknown length
21162 21161: contig of 1074 bp in length
21261 21261: gap of unknown length

21262 22935: contig of 1674 bp in length
22936 23035: gap of unknown length
23036 24725: contig of 1691 bp in length
24727 24826: gap of unknown length
24827 25984: contig of 1158 bp in length
25985 26084: gap of unknown length
26085 28117: contig of 2033 bp in length
28118 28217: gap of unknown length
28218 28226: contig of 1009 bp in length
29227 29326: gap of unknown length
29327 31330: contig of 2004 bp in length
31331 31430: gap of unknown length
31431 33008: contig of 1578 bp in length
33009 33108: gap of unknown length
33109 34893: contig of 1785 bp in length
34894 34993: gap of unknown length
34994 36418: contig of 1425 bp in length
36419 36518: gap of unknown length
36519 38755: contig of 2237 bp in length
38756 38855: gap of unknown length
38856 40715: contig of 1860 bp in length
40716 40815: gap of unknown length
40816 42591: contig of 1776 bp in length
42592 42691: gap of unknown length
42692 44432: contig of 1741 bp in length
44433 44532: gap of unknown length
44534 46670: contig of 2138 bp in length
46671 46770: gap of unknown length
46771 48171: contig of 1401 bp in length
48172 48271: gap of unknown length
48272 50293: contig of 2022 bp in length
50294 50393: gap of unknown length
50394 52099: contig of 1706 bp in length
52100 52199: gap of unknown length
52200 54742: contig of 2543 bp in length
54743 54842: gap of unknown length
54843 57839: contig of 2997 bp in length
57840 57939: gap of unknown length
57940 59989: contig of 2050 bp in length
59990 60089: gap of unknown length
60090 61699: contig of 1510 bp in length
61699 61600: gap of unknown length
61700 63235: contig of 1536 bp in length
63236 63335: gap of unknown length
63336 65866: contig of 2361 bp in length
65867 65796: gap of unknown length
65797 68230: contig of 2434 bp in length
68231 68330: gap of unknown length
68331 70088: contig of 1758 bp in length
70089 70188: gap of unknown length
70189 72866: contig of 2678 bp in length
72867 72966: gap of unknown length
72967 76485: contig of 3519 bp in length
76486 76585: gap of unknown length
76586 78904: contig of 2319 bp in length
78905 79004: gap of unknown length
79006 81753: contig of 2759 bp in length
81754 81853: gap of unknown length
81854 85381: contig of 3518 bp in length
85382 85481: gap of unknown length
85482 88489: contig of 3008 bp in length
88490 88589: gap of unknown length
88590 91317: contig of 2728 bp in length
91318 91417: gap of unknown length
91418 93443: contig of 2026 bp in length
93444 95582: contig of 2039 bp in length
95583 95682: gap of unknown length
95683 99140: contig of 3458 bp in length
99141 99240: gap of unknown length
99241 101817: contig of 2577 bp in length
101818 101917: gap of unknown length
101918 105395: contig of 3478 bp in length

Query Match 80.8%; Score 20.2; DB 5; Length 175;
 Best Local Similarity 88.0%; Pred. No. 52;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 140 GCTTCTTGGCCAGAGCTCTGCCA 164

RESULT 11
 AF249134 175 bp DNA linear VRT 17-JAN-2001
 LOCUS Mantisidactylus cf. ulcerosus rhodopsin gene, exon 4 and partial cds.
 DEFINITION
 ACCESSION AF249134
 VERSION AF249134.1 GI:12247197
 KEYWORDS
 SOURCE Mantisidactylus cf. ulcerosus.
 ORGANISM Mantisidactylus cf. ulcerosus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rhacophoridae;

REFERENCE 1 (bases 1 to 175)
 Bossuyt, F. and Milinkovitch, M.C.
 Convergent Adaptive Radiations in Madagascan and Asian Rain Forest Frogs
 TITLE

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 175)
 Bossuyt, F. and Milinkovitch, M.C.
 Direct Submission
 TITLE
 AUTHORS
 JOURNAL Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of Molecular Biology and Medicine, rue Jeener and Brachet 12, Gosselies B-6041, Belgium

FEATURES
 Source Location/Qualifiers

1..175
 /organism="Mantisidactylus cf. ulcerosus"
 /db_xref="taxon:129014"
 <1..>175
 /product="rhodopsin"

CDS
 <1..>175
 /codon_start=2
 /product="rhodopsin"
 /protein_id="AAG49777.1"
 /db_xref="GI:12247198"
 /translation="AEKEVTRMVIVVFFLICWVYAVVAFYIFTHGSEFGPIFMT
 VPAFAKSSAIYNP"
 <1..>175
 exon

BASE COUNT 36 a 50 c 36 g 53 t
 ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 175;
 Best Local Similarity 88.0%; Pred. No. 52;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 140 GCTTCTTGGCCAGAGCTCTGCCA 164

RESULT 12
 AF249136 175 bp DNA linear VRT 17-JAN-2001
 LOCUS Boophis xerophilus rhodopsin gene, exon 4 and partial cds.
 DEFINITION
 ACCESSION AF249136
 VERSION AF249136.1 GI:12247201
 KEYWORDS

SOURCE
 ORGANISM Boophis xerophilus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rhacophoridae;
 Boophis.

REFERENCE 1 (bases 1 to 175)
 Bossuyt, F. and Milinkovitch, M.C.
 Convergent Adaptive Radiations in Madagascan and Asian Rain Forest Frogs
 TITLE
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 175)
 Bossuyt, F. and Milinkovitch, M.C.
 Direct Submission
 TITLE
 AUTHORS
 JOURNAL Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of Molecular Biology and Medicine, rue Jeener and Brachet 12, Gosselies B-6041, Belgium

FEATURES
 source Location/Qualifiers

1..175
 /organism="Boophis xerophilus"
 /db_xref="taxon:128996"
 <1..>175
 /product="rhodopsin"

CDS
 <1..>175
 /codon_start=2
 /product="rhodopsin"
 /protein_id="AAG49779.1"
 /db_xref="GI:12247202"
 /translation="AEKEVTRMVIVVFFLICWVYAVVAFYIFTHGSEFGPIFMT
 VPAFAKSSAIYNP"
 <1..>175
 exon

BASE COUNT 34 a 55 c 34 g 52 t
 ORIGIN

Query Match 80.8%; Score 20.2; DB 5; Length 175;
 Best Local Similarity 88.0%; Pred. No. 52;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 140 GCTTCTTGGCCAGAGCTCTGCCA 164

RESULT 13
 AF249137 175 bp DNA linear VRT 17-JAN-2001
 LOCUS Boophis tephraeomystax rhodopsin gene, exon 4 and partial cds.
 DEFINITION
 ACCESSION AF249137
 VERSION AF249137.1 GI:12247203
 KEYWORDS
 SOURCE Boophis tephraeomystax.
 ORGANISM Boophis tephraeomystax.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rhacophoridae;
 Boophis.

REFERENCE 1 (bases 1 to 175)
 Bossuyt, F. and Milinkovitch, M.C.
 Convergent Adaptive Radiations in Madagascan and Asian Rain Forest Frogs
 TITLE
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 175)
 Bossuyt, F. and Milinkovitch, M.C.
 Direct Submission
 TITLE
 AUTHORS
 JOURNAL Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of Molecular Biology and Medicine, rue Jeener and Brachet 12, Gosselies B-6041, Belgium

FEATURES
 source Location/Qualifiers

1..175
 /organism="Boophis tephraeomystax"
 /db_xref="taxon:68440"
 <1..>175
 /product="rhodopsin"

```

/codon_start=2
/product="rhodopsin"
/protein_id="AAG49780.1"
/db_xref="GI:12247204"
/translation="AEKEVTRMVIIMVFFLLICWPRYAVAFYIFTHQSEFGPIFMT
VPAFFAKSSAIYNP"
<1..>175
/number=4
BASE COUNT      35 a      56 c      33 g      51 t
ORIGIN
Query Match      80.8%; Score 20.2; DB 5; Length 175;
Best Local Similarity 88.0%; Pred. No. 52;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCTTCTTGCTGAGAGCTCTGCCA 25
|||||
Db 140 GCTTCTTGCCAGAGCTCTGCCA 164

RESULT 14
AF249138      175 bp      DNA      linear      VRT 17-JAN-2001
LOCUS      Laliostoma labrosus rhodopsin gene, exon 4 and partial cds.
DEFINITION
ACCESSION      AF249138
VERSION
KEYWORDS
SOURCE
ORGANISM      Laliostoma labrosus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae;
Laliostoma.
1 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Convergent Adaptive Radiations in Madagascar and Asian Ranid Frogs
Reveal Co-variation between Larval and Adult Traits
Unpublished
2 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Direct Submission
Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
Molecular Biology and Medicine, rue Jeener and Brachet 12,
Gosselies B-6041, Belgium
Location/Qualifiers
1..175
/organism="Laliostoma labrosus"
/db_xref="taxon:129011"
<1..>175
/product="rhodopsin"
<1..>175
/codon_start=2
/product="rhodopsin"
/protein_id="AAG49780.1"
/db_xref="GI:12247204"
/translation="AEKEVTRMVIIMVFFLLICWPRYAVAFYIFTHQSEFGPIFMT
VPAFFAKSSAIYNP"
<1..>175
/number=4
BASE COUNT      35 a      56 c      36 g      49 t
ORIGIN
Query Match      80.8%; Score 20.2; DB 5; Length 175;
Best Local Similarity 88.0%; Pred. No. 52;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCTTCTTGCTGAGAGCTCTGCCA 25
|||||
Db 140 GCTTCTTGCCAGAGCTCTGCCA 164

RESULT 15
AF249139      175 bp      DNA      linear      VRT 17-JAN-2001
LOCUS

```

```

DEFINITION      Fejervarya syhadrensis rhodopsin gene, exon 4 and partial cds.
ACCESSION      AF249139
VERSION      AF249139.1 GI:12247207
KEYWORDS
SOURCE
ORGANISM      Fejervarya syhadrensis.
Fejervarya syhadrensis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae;
Fejervarya.
1 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Convergent Adaptive Radiations in Madagascar and Asian Ranid Frogs
Reveal Co-variation between Larval and Adult Traits
Unpublished
2 (bases 1 to 175)
Bossuyt, F. and Milinkovitch, M.C.
Direct Submission
Submitted (27-MAR-2000) Unit of Evolutionary Genetics, Institute of
Molecular Biology and Medicine, rue Jeener and Brachet 12,
Gosselies B-6041, Belgium
Location/Qualifiers
1..175
/organism="Fejervarya syhadrensis"
/db_xref="taxon:128998"
<1..>175
/product="rhodopsin"
<1..>175
/codon_start=2
/product="rhodopsin"
/protein_id="AAG49780.1"
/db_xref="GI:12247208"
/translation="AEKEVTRMVIIMVFFLLICWPRYAVAFYIFTHQSEFGPIFMT
VPAFFAKSSAIYNP"
<1..>175
/number=4
BASE COUNT      34 a      50 c      35 g      54 t      2 others
ORIGIN
Query Match      80.8%; Score 20.2; DB 5; Length 175;
Best Local Similarity 88.0%; Pred. No. 52;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCTTCTTGCTGAGAGCTCTGCCA 25
|||||
Db 140 GCTTCTTGCCAGAGCTCTGCCA 164

Search completed: March 17, 2003, 11:31:57
Job time : 256.108 secs

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:42:12 ; Search time 143.28 Seconds

(without alignments)
392.938 Million cell updates/sec

Title: US-09-836-439-5

Perfect score: 25

Sequence: 1 gcttcttctgctgagctcttcca 25

Scoring table: IDENTITY-NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_101002:*

1: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1980.DAT:*

2: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1981.DAT:*

3: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1982.DAT:*

4: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1983.DAT:*

5: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1984.DAT:*

6: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1985.DAT:*

7: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1986.DAT:*

8: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1987.DAT:*

9: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1988.DAT:*

10: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1989.DAT:*

11: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1990.DAT:*

12: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1991.DAT:*

13: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1992.DAT:*

14: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1993.DAT:*

15: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1994.DAT:*

16: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1995.DAT:*

17: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1996.DAT:*

18: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1997.DAT:*

19: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1998.DAT:*

20: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA1999.DAT:*

21: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA2000.DAT:*

22: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA2001A.DAT:*

23: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA2001B.DAT:*

24: /SIDS2/gcgdata/geneq/geneq-n-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.2	72.8	807	21	AAAC48879
C 2	18.2	72.8	906	23	AAAT77337
C 3	18.2	72.8	3254	22	AAAC85295
C 4	18.2	72.8	6912	22	AAAC36854
C 5	17.8	71.2	702	23	ABL26659
C 6	17.8	71.2	756	21	AAH51547
C 7	17.8	71.2	1000	21	AAH51581
C 8	17.8	71.2	1000	22	AAH51581
C 9	17.8	71.2	2573	19	AAV59153

10	17.8	71.2	3153	23	ABL26658	Drosophila melanog
11	17.8	71.2	3870	23	ABL26632	Drosophila melanog
12	17.8	71.2	8002	24	AAAD31629	Arabidopsis thaliana
C 13	17.6	70.4	223	21	AAAC07949	Human secreted pro
C 14	17.6	70.4	342	24	ABL69358	Prostate cancer re
C 15	17.6	70.4	384	23	ABV11069	Human prostate exp
C 16	17.6	70.4	403	24	ABN26150	Human prostate exp
C 17	17.6	70.4	437	23	ABV32215	Human prostate exp
C 18	17.6	70.4	437	23	ABV41146	Human prostate exp
C 19	17.6	70.4	437	23	ABV41150	Human prostate exp
C 20	17.6	70.4	437	23	ABV41150	Human prostate exp
C 21	17.6	70.4	745	23	AAZ41246	Human normal ovar
C 22	17.6	70.4	749	23	AAZ41246	Human normal ovar
C 23	17.6	70.4	761	20	AAZ42214	DNA encoding novel
C 24	17.6	70.4	779	20	AAZ42214	Human normal blad
C 25	17.6	70.4	5252	22	AAZ42214	Human normal blad
C 26	17.6	70.4	23899	23	AAZ42214	Rat CARD-6 CDNA.
C 27	17.6	69.6	2849	24	ABL53927	Drosophila melanog
C 28	17.2	68.8	570	22	AAZ42214	Human dihydroorot
C 29	17.2	68.8	687	22	AAZ42214	Human breast cance
C 30	17.2	68.8	712	22	AAZ42214	Aspergillus oryzae
C 31	17.2	68.8	712	22	AAZ42214	Human breast cance
C 32	17.2	68.8	811	22	AAZ42214	Human breast cance
C 33	17.2	68.8	850	21	AAZ42214	Human breast cance
C 34	17.2	68.8	904	21	AAZ42214	Partial sequence M
C 35	17.2	68.8	1512	22	AAZ42214	Partial sequence M
C 36	17.2	68.8	1578	22	AAZ42214	Human cDNA encodin
C 37	17.2	68.8	1676	21	AAZ42214	Human cDNA encodin
C 38	17.2	68.8	1739	22	AAZ42214	Human ORFX ORP1428
C 39	17.2	68.8	1828	22	AAH15348	Human full-length
C 40	17.2	68.8	1934	22	AAH15348	Human cDNA sequenc
C 41	17.2	68.8	1945	22	AAH15348	Human protein havt
C 42	17.2	68.8	1945	24	ABL0687	Human secreted pro
C 43	17.2	68.8	2015	22	AAH15348	Human polynucleot
C 44	17.2	68.8	2015	22	AAH15348	Nucleotide sequenc
C 45	17.2	68.8	2015	22	AAH15348	Human secreted pro

ALIGNMENTS

RESULT 1	AAAC48879/c	AAAC48879 standard; DNA: 807 BP.
ID	AAAC48879	
AC	AAAC48879	
XX		
XX		
18-OCT-2000	(first entry)	
DT		
XX		
XX		
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 59111.	
KW	Hybridisation assay; genetic mapping; gene expression control;	
KW	protein identification; signal transduction pathway;	
KW	metabolic pathway; promoter; termination sequence; ss.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
06-SEP-2000.		
PD		
XX		
XX		
25-FEB-2000;	2000EP-0301439.	
XX		
25-FEB-1999;	9905-0121825.	
PR	05-MAR-1999;	9905-0123180.
PR	09-MAR-1999;	9905-0123548.
PR	23-MAR-1999;	9905-0125788.
PR	25-MAR-1999;	9905-0126264.
PR	29-MAR-1999;	9905-0126785.
PR	01-APR-1999;	9905-0127462.
PR	06-APR-1999;	9905-0128234.
PR	08-APR-1999;	9905-0128714.
PR	16-APR-1999;	9905-0129645.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132487.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140683.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142350.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145912.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 29-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 990S-0159330.
 PR 14-OCT-1999; 990S-0159331.
 PR 14-OCT-1999; 990S-0159637.
 PR 14-OCT-1999; 990S-0159638.
 PR 14-OCT-1999; 990S-0159638.
 PR 21-OCT-1999; 990S-0160741.
 PR 21-OCT-1999; 990S-0160741.
 PR 21-OCT-1999; 990S-0160768.
 PR 21-OCT-1999; 990S-0160770.
 PR 21-OCT-1999; 990S-0160814.
 PR 21-OCT-1999; 990S-0160815.
 PR 22-OCT-1999; 990S-0160980.
 PR 22-OCT-1999; 990S-0160981.
 PR 25-OCT-1999; 990S-0161404.
 PR 25-OCT-1999; 990S-0161405.
 PR 25-OCT-1999; 990S-0161406.
 PR 26-OCT-1999; 990S-0161359.
 PR 26-OCT-1999; 990S-0161360.
 PR 26-OCT-1999; 990S-0161361.
 PR 28-OCT-1999; 990S-0161920.
 PR 28-OCT-1999; 990S-0161992.
 PR 28-OCT-1999; 990S-0161993.
 PR 29-OCT-1999; 990S-0162142.

Query Match 72.8%; Score 18.2; DB 21; Length 807;
 Best Local Similarity 87.0%; Pred. No. 1.4e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 CTTTCTTTCCTGAGAGCTCTCC 24
 DB 401 CTTGTTTCTGCTGAGATCTCTCC 379

RESULT 2

AA577337/c
 ID AA577337 standard; CDNA: 906 BP.

AC AA577337;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #13141.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG13150.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

PS Claim 1; SEQ ID No 13141; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AA564197-AA594564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

Sequence 906 BP; 243 A; 218 C; 269 G; 176 T; 0 other;
 Query Match 72.8%; Score 18.2; DB 23; Length 906;
 Best Local Similarity 87.0%; Pred. No. 1.4e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCTTCTTTCCTGAGAGCTCTTC 23
 DB 677 GCTTCTTTCCTGAGATCTCTTC 655

RESULT 3

AA585295
 ID AA585295 standard; CDNA: 3254 BP.

AC AA585295;

DT 29-MAR-2001 (first entry)

DE Mouse SPANK CDNA.

XX SPANK: SAM; sterile alpha motif; PAPP; insulin resistance;

KW poly adenosine diphosphate-ribose polymerase; catalytic domain;

KW ANK; ankyrin repeat; cytosol; insulin-responsive aminopeptidase;

KW IRAP; GLUT4; adipocyte; insulin signalling pathway; hyperlipidaemia;

KW glucose intolerance; atherosclerotic disease; atherosclerosis;

KW obesity; cardiac insufficiency; coronary insufficiency; stroke;

KW high blood pressure; non-insulin dependent diabetes; hypertension;

KW hyperuricemia; Syndrome X; muscular dystrophy; muscle atrophy; ds.

XX Mus musculus.

OS Mus musculus.

PH Key

FT CDS

FT 308..3254

FT /tag= a

FT /product= "Murine SPANK"

FT /transl_except= (pos:1754,aa:Gly)

FT /note= "This codon has an apparent 2 nucleotide

FT deletion, which alters the reading frame"

FT /transl_except= (pos:2145,aa:DIOGKRLIG)

FT /note= "The nucleotides encoding residues 614 to 623

FT are absent, which alters the reading frame"

FT /transl_except= (pos:2602..2606,aa:Met)

FT /note= "This codon has an apparent 2 nucleotide

FT insertion, which alters the reading frame"

WO200077225-A1.

21-DEC-2000.

09-JUN-2000; 2000MO-US15926.

PR 11-JUN-1999; 99US-0138957.
XX (WHEED) WHITEHEAD INST BIOMEDICAL RES.
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Chi N, Lodish HF;
XX
DR WPI, 2001-091404/10.
DR P-PSDB; AAB47023.
XX
PT New insulin signalling protein SPANK, useful for reducing body mass,
PT glucose intolerance or insulin resistance and for preventing or
PT treating obesity-related and muscle-related diseases -
XX
PS Claim 3; Fig 5A; 65pp: English.
XX
CC This sequence represents the mouse SPANK cDNA. The SPANK
CC protein comprises 3 domains:
CC (a) a SAM (sterile alpha motif) domain;
CC (b) a PAM (poly adenosine diphosphate-ribose polymerase) catalytic
CC domain; and
CC (c) an ANK domain composed of ankyrin repeats.
CC SPANK is a cytosolic protein which can poly(ADP-ribosyl)ate itself.
CC SPANK binds insulin-responsive aminopeptidase (IRAP) and modulates
CC translocation of GLUT4 in the perinuclear region of adipocytes. It
CC is an effector in the insulin signalling pathway in eukaryotic cells.
CC SPANK is useful for reducing body mass, reducing glucose
CC intolerance or insulin resistance, for preventing or treating
CC obesity-related diseases or disorders, such as obesity, cardiac
CC insufficiency, coronary insufficiency, stroke, hypertension,
CC atherosclerosis, hyperlipidaemia, high blood pressure, non-insulin
CC dependent diabetes, hyperlipidaemia, hyperuricaemia and syndrome X and is
CC also useful for preventing or treating muscle-related diseases or
CC disorders, such as muscular dystrophy, muscle atrophy and muscle
CC fatigue. Antibodies immunospecific for SPANK are useful for detecting
CC the presence of SPANK polypeptide in a biological sample.
XX
SQ Sequence 3254 BP; 871 A; 746 C; 863 G; 774 T; 0 other;
XX
Query Match 72.8%; Score 18.2; DB 22; Length 3254;
Best Local Similarity 87.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 3 TTCTTGTGCTGAGACTCTTCCA 25
II |||||||||
Db 2930 TTATTTGCTGAGAACTCTTCCA 2952
XX
RESULT 4
AAS36854
ID AAS36854 standard; DNA; 6912 BP.
XX
AC AAS36854;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human cardiovascular system antigen genomic DNA SEQ ID No 2354.
XX
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antihemmatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX
OS Homo sapiens.
XX
PN WO200155321-A2.

XX 02-AUG-2001.
PD 17-JAN-2001; 2001WO-US01340.
XX
PR 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0235837.

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM.
XX

DR WPI; 2001-451930/48.
XX
PT New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
PS Claim 1; SEQ ID NO 2354; 674pp; English.
XX
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Query Match 72.8%; Score 18.2; DB 22; Length 6912;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 3 TTCTTGTGCTGAGAGCTCTTCCA 25
|||||
Db 647 TTCTTGTGCTGATGACGGTTCCA 669

RESULT 5
ABL26659
ID ABL26659 standard; DNA; 702 BP.
XX
AC ABL26659;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31450.
XX
KM Drosophila: developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT Interactions -

XX Claim 1; SEQ ID NO 31450; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB130511), expressed DNA sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 702 BP; 152 A; 190 C; 158 G; 202 T; 0 other;

SO Query Match 71.2%; Score 17.8; DB 23; Length 702;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 TCTTTGCTGAGAGCTCTTCCA 25
||||| ||||| ||||| |||||
Db 475 TCTTTGCTGAGAGCTCTTCCA 495

RESULT 6
AAH51547/c
ID AAH51547 standard; DNA; 756 BP.
XX AAH51547;
AC
XX 29-AUG-2001 (first entry)
DE Human MGSTII related DNA containing a biallelic polymorphism SEQ ID 438.
XX Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII; microsomal glutathione S-transferase II; malate decarboxylase enzyme; DMEL; MEL; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5; flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5; dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype; phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype; uridine diphosphate glucuronosyl transferase; UGT2; asthma; hepatotoxicity; zileuton; ds.
XX Homo sapiens.
XX WO200058508-A2.
XX 05-OCT-2000.
XX 24-MAR-2000; 2000WO-IB00403.
XX 25-MAR-1999; 99US-0126269.
XX 30-APR-1999; 99US-0131961.
XX (GEST) GENSET.
XX Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX WPI; 2000-638353/61.
XX Polynucleotides comprising sequences from malate decarboxylase PT enzyme-related biallelic markers used for genotyping -
XX Claim 13; Page 616; 673bp; English.
XX Sequences AAH51110-AAH51593 represent human DNA fragments which contain biallelic markers. The sequences are related to various human genes including microsomal glutathione S-transferase II (MGSTII), malate decarboxylase enzyme (DMEL/MEL), cytochrome P450, glutathione reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),

CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine diphosphate glucuronosyl transferases (UGT2). Each of these sequences contains a biallelic marker/polymorphism, which is represented in the sequence as a degenerate/undefined base. The genes to which the biallelic marker containing sequences are related are involved in drug metabolism.

CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the MGSTII gene and four alternative MGSTII cDNA sequences. AAB62905-AAH62906 are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in an example for the amplification of human genomic DNA fragments. The invention includes a method of genotyping comprising determining the identity of a nucleotide at a DME- or MGSTII-related biallelic marker in a biological sample. The method is used to determine the frequency in population of an allele of a DME- or MGST-II related biallelic marker and to select an individual for inclusion in a clinical trial of a drug treatment. The method is also used to detect association between allele CC and phenotype, and to detect association between haplotype and phenotype. The polynucleotides are used, in hybridization assays, sequencing assays CC or allele specific amplification assays. The method can be used to determine whether an individual suffers or is at risk of developing CC asthma or is at risk of developing hepatotoxicity on treatment with CC zileuton.

XX Sequence 756 BP; 224 A; 140 C; 177 G; 213 T; 2 other;

SO Query Match 71.2%; Score 17.8; DB 21; Length 756;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTTCTTGTGAGAGCTCTTC 23
||||| ||||| ||||| |||||
Db 614 TTTCTTGTGAGAGCTCTTC 594

RESULT 7
AAH51581/c
ID AAH51581 standard; DNA; 1000 BP.
XX AAH51581;
AC
XX 29-AUG-2001 (first entry)
DE Human MGSTII related DNA containing a biallelic polymorphism SEQ ID 472.
XX Human; biallelic marker; single nucleotide polymorphism; SNP; MGSTII; microsomal glutathione S-transferase II; malate decarboxylase enzyme; DMEL; MEL; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5; flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5; dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype; phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype; uridine diphosphate glucuronosyl transferase; UGT2; asthma; hepatotoxicity; zileuton; ds.
XX Homo sapiens.
XX WO200058508-A2.
XX 05-OCT-2000.
XX 24-MAR-2000; 2000WO-IB00403.
XX 25-MAR-1999; 99US-0126269.
XX 30-APR-1999; 99US-0131961.
XX (GEST) GENSET.
XX Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX WPI; 2000-638353/61.
XX Polynucleotides comprising sequences from malate decarboxylase PT enzyme-related biallelic markers used for genotyping -
XX

CC Insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 3870 BP; 961 A; 917 C; 886 G; 1106 T; 0 other;
XX
Query Match 71.2%; Score 17.8; DB 23; Length 3870;
Best Local Similarity 90.5%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 5 TCTTTGCTGAGAGCTCTTCCA 25
DB 103 TCTTCTGCTGAGAGCTCTTCCA 123
XX
RESULT 12
ID AAB31629 standard; DNA; 8002 BP.
XX
AC AAD31629;
XX
DT 18-JUN-2002 (first entry)
XX
DE Arabidopsis thaliana GT1209 gene.
XX
KM GT1209 gene; herbicide; plant tissue; cell; seedling growth; development;
XX ds.
XX Arabidopsis thaliana.
XX
OS Arabidopsis thaliana.
XX
PN W0200212273-A2.
XX
PD 14-FEB-2002.
XX
PF 01-AUG-2001; 2001WO-EP08910.
XX
PR 03-AUG-2000; 2000US-222779P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Levin JZ, Wegrich Glover L, Budziszewski CJ;
XX
DR MPI; 2002-241730/29.
XX
PT New polypeptide having GT1209, GT1354 or GT0946 activity, obtained from
PT Arabidopsis, useful as herbicide targets in screening assays to
PT identify the inhibitors or potential herbicides
XX
PS Disclosure: Page 72-74; 82pp; English.
XX
CC The invention relates to a polypeptide having GT1209, GT1354 or GT0946
CC activity obtained from Arabidopsis. The invention may also be applied to
CC the development of herbicide tolerant plants, plant tissues, plant seeds
CC and plant cells. The polypeptide is useful as herbicide targets in
CC screening assays to identify potential herbicides and inhibitors of
CC GT1209, GT1354 or GT0946 activity. A compound having herbicidal activity
CC is useful for suppressing the growth of a plant. The newly discovered
CC GT1209, GT1354 or GT0946 genes are essential for seedling growth and
CC development. The present sequence is Arabidopsis thaliana GT1209 gene.
XX
SQ Sequence 8002 BP; 2119 A; 1489 C; 1502 G; 2892 T; 0 other;
XX
Query Match 71.2%; Score 17.8; DB 24; Length 8002;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 CTTTCTTGTGAGAGCTCTT 22
DB 6994 CTTTCTTGTGAGAGCTCTT 7014
XX

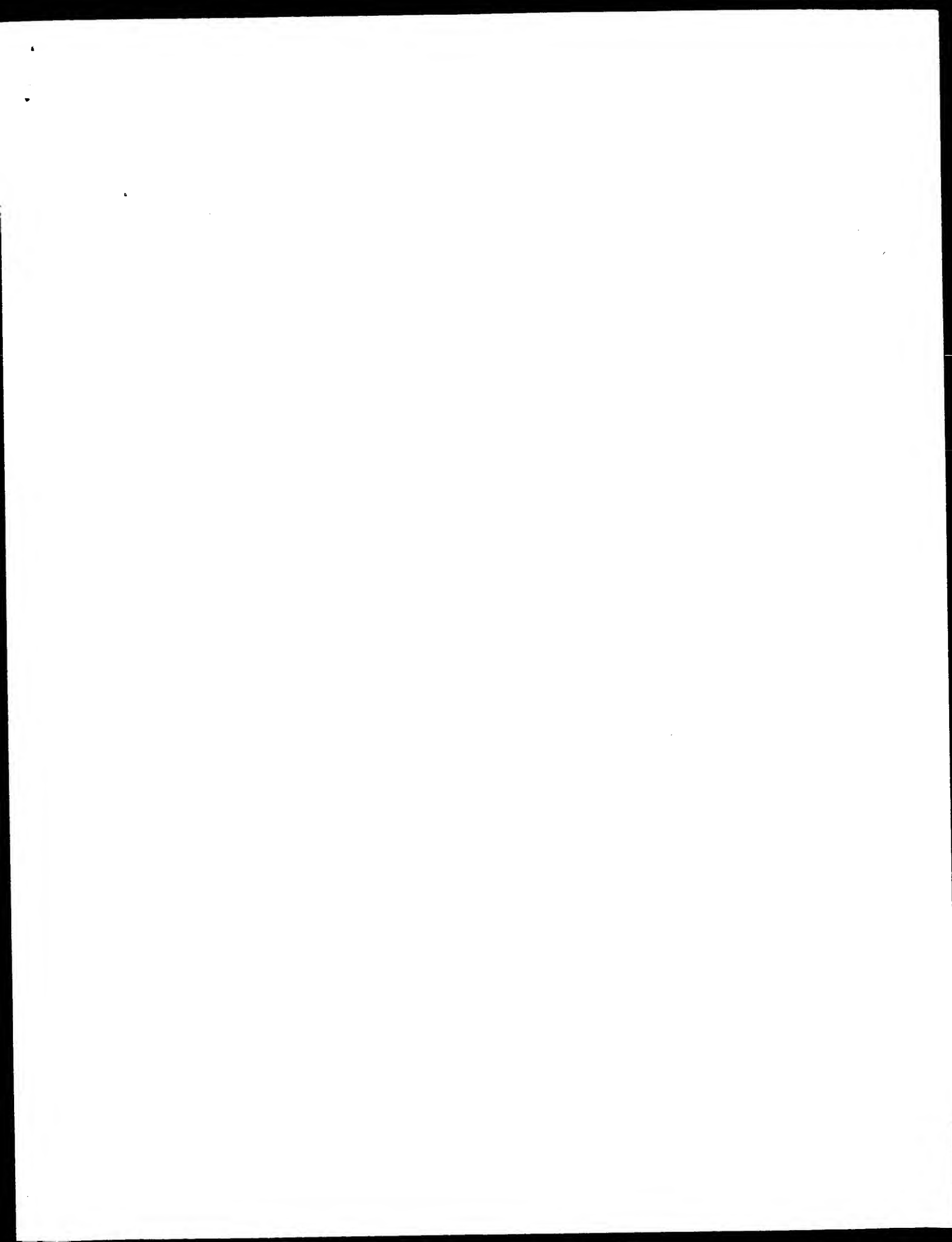
RESULT 13
ID AAC07949/C
AAC07949 standard; cDNA; 223 BP.
XX
AC AAC07949;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 12024.
XX
KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR MPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 12024; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 223 BP; 77 A; 38 C; 53 G; 54 T; 1 other;
XX
Query Match 70.4%; Score 17.6; DB 21; Length 223;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 GCTTCTTGTGAGAGCTCTTCC 24
DB 152 GCTTCTTGTGAGAGATTTCTTCC 129
XX
RESULT 14
ID ABL69358
ABL69358 standard; DNA; 342 BP.
XX
AC ABL69358;
XX
DT 15-MAY-2002 (first entry)
XX
DE Prostate cancer related gene sequence SEQ ID NO: 7695.
XX

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200194629-A2.
 XX
 PD 13-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-US10838.
 XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-233133P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234009P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 27-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX
 PA (AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 DR WPI; 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -
 XX
 XX Claim 1; SEQ ID 7695; 44pp; English.
 XX
 CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)

CC comprises a sequence (S) selected from 8447 sequences (given in AB161664
 CC to AB170110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilm's tumour.
 XX
 SQ Sequence 342 BP; 36 A; 109 C; 92 G; 105 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 24; Length 342;
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GCTTCTTGTGCTGAGGCTCC 24
 Db 67 GCTTCTTGTGCTGAGGCTCC 90
 RESULT 15
 ID ABY11069 standard; cDNA; 384 BP.
 XX
 XX ABY11069;
 XX
 DT 13-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 11060.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 PI
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 XX Claim 1; Page 1792; 11750pp; English.
 XX
 PS The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABY00010-ABY62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 384 BP; 139 A; 87 C; 73 G; 85 T; 0 other;
Query Match .70.4%; Score 17.6; DB 23; Length 384;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 2 CTTCTTGGGAGAGCTTTCA 25
||||||| ||||| |||
Db 374 CTTCTTGGGAGAGCTTTCA 351

Search completed: March 17, 2003, 10:50:50
Job time : 146.446 secs



GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 883.737 Seconds

(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-5

Perfect score: 25

Sequence: 1 gcttcttgcgtgagcgtcttcca 25

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database :
EST :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estcl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vyt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23.4	93.6	402	12	BE950666
2	23.4	93.6	495	13	BE950666
3	23.4	93.6	628	13	BE950666
4	23.4	93.6	636	12	BE950666
5	23.4	93.6	649	12	BE950666
6	23.4	93.6	651	13	BE950666

7	23.4	93.6	656	13	BE950666
8	23.4	93.6	663	13	BE950666
9	23.4	93.6	672	12	BE950666
10	23.4	93.6	676	13	BE950666
11	23.4	93.6	681	13	BE950666
12	23.4	93.6	682	13	BE950666
13	23.4	93.6	684	12	BE950666
14	23.4	93.6	686	12	BE950666
15	23.4	93.6	687	12	BE950666
16	23.4	93.6	687	14	BE950666
17	23.4	93.6	697	13	BE950666
18	23.4	93.6	717	13	BE950666
19	23.4	93.6	718	13	BE950666
20	23.4	93.6	720	13	BE950666
21	23.4	93.6	720	13	BE950666
22	23.4	93.6	724	12	BE950666
23	23.4	93.6	744	12	BE950666
24	23.4	93.6	746	12	BE950666
25	23.4	93.6	759	12	BE950666
26	23.4	93.6	761	12	BE950666
27	23.4	93.6	773	13	BE950666
28	23.4	93.6	774	12	BE950666
29	23.4	93.6	783	13	BE950666
30	23.4	93.6	790	13	BE950666
31	23.4	93.6	802	13	BE950666
32	23.4	93.6	806	12	BE950666
33	23.4	93.6	825	13	BE950666
34	23.4	93.6	832	13	BE950666
35	23.4	93.6	837	12	BE950666
36	23.4	93.6	851	13	BE950666
37	23.4	93.6	854	13	BE950666
38	23.4	93.6	873	13	BE950666
39	23.4	93.6	879	13	BE950666
40	23.4	93.6	880	14	BE950666
41	23.4	93.6	903	12	BE950666
42	23.4	93.6	909	14	BE950666
43	23.4	93.6	914	14	BE950666
44	23.4	93.6	917	12	BE950666
45	23.4	93.6	918	13	BE950666

ALIGNMENTS

RESULT 1
LOCUS BE950666/c
DEFINITION UI-M-CE0-aza-d-06-0-UI-s1 NIH-BMAP Ret3 Mus musculus cDNA clone
ACCESSION BE950666
VERSION BE950666.1 GI:10589332
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 402)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@nhi.nih.gov
The sequence contained an oligo-dT track that was present in the strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NciI site

and the oligo-dT track served to verify it as a clone from the retina tissue cDNA library. Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements. The tissue for this library was contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine. Seq primer: M13 Forward

FEATURES

SOURCE

Location/Qualifiers
1. 402
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CEO-aza-d-06-0-UI"
/clone_lib="NIH_BMAP_Ret3"
/dev_stage="6 weeks"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_Ret3 library is derived from mouse retina tissue. For a detailed description of the library from which this clone was derived, please visit our web site at bradnest.eng.uiowa.edu. The tissue for this library was contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine.
TAG_LIB="NIH_BMAP_Ret3"
TAG_TISSUE="adult-retina"
TAG_SEQ="GTCAGCGCAC"
BASE COUNT 90 a 86 c 126 g 100 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 12; Length 402;
Best Local Similarity 96.0%; Pred. No. 19;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGTGAGAGCTCTTCCA 25
|||||
Db 334 GCTTCTTGTGAGAGCTCTTCCA 310

RESULT 2 495 bp mRNA linear EST 20-SEP-2001
LOCUS B1737699 603358627F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5365638 5',
DEFINITION mRNA sequence.
ACCESSION B1737699
VERSION B1737699.1 GI:15714712
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
1 (bases 1 to 495)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM1930 row: f column: 07
High quality sequence stop: 489.
Location/Qualifiers
1. 495
/organism="Mus musculus"
/db_xref="taxon:10090"

FEATURES

SOURCE

1. 495
/organism="Mus musculus"
/db_xref="taxon:10090"

BASE COUNT 105 a 159 c 111 g 120 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 495;
Best Local Similarity 96.0%; Pred. No. 21;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGTGAGAGCTCTTCCA 25
|||||
Db 270 GCTTCTTGTGAGAGCTCTTCCA 294

RESULT 3 628 bp mRNA linear EST 20-SEP-2001
LOCUS B1734160 603351436F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5358776 5',
DEFINITION mRNA sequence.
ACCESSION B1734160
VERSION B1734160.1 GI:15711173
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
1 (bases 1 to 628)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM1912 row: h column: 09
High quality sequence start: 4
High quality sequence stop: 627.
Location/Qualifiers
1. 628
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5358776"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."

BASE COUNT 159 a 180 c 152 g 137 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 628;
Best Local Similarity 96.0%; Pred. No. 23;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGTGAGAGCTCTTCCA 25
|||||
Db 45 GCTTCTTGTGAGAGCTCTTCCA 69

RESULT 4
 B1736587 636 bp mRNA linear EST 20-SEP-2001
 LOCUS 603361053F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:536804 5',
 DEFINITION mRNA sequence.
 ACCESSION B1736587
 VERSION B1736587.1 GI:15713600
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 636)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1936 row: 3 column: 13
 High quality sequence stop: 632.
 Location/Qualifiers
 1..636
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:536804"
 /clone_1ib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 145 a 208 c 147 g 136 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 636;
 Best Local Similarity 96.0%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGGCTGAGAGCTCTTCCA 25
 ||||||||| |||||||||
 Db 179 GCTTCTTGGCTGAGAGCTCTTCCA 203

RESULT 5
 BG403992 649 bp mRNA linear EST 12-MAR-2001
 LOCUS 602419859F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:452696 5',
 DEFINITION mRNA sequence.
 ACCESSION BG403992
 VERSION BG403992.1 GI:13297440
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 649)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10435 row: e column: 15
 High quality sequence stop: 644.
 Location/Qualifiers
 1..649
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:452696"
 /clone_1ib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 164 a 203 c 140 g 142 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 12; Length 649;
 Best Local Similarity 96.0%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGGCTGAGAGCTCTTCCA 25
 ||||||||| |||||||||
 Db 108 GCTTCTTGGCTGAGAGCTCTTCCA 132

RESULT 6
 B1730233 651 bp mRNA linear EST 20-SEP-2001
 LOCUS 60335020F1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:5357472 5',
 DEFINITION mRNA sequence.
 ACCESSION B1730233
 VERSION B1730233.1 GI:15707246
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 651)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1909 row: b column: 01
 High quality sequence stop: 539.
 Location/Qualifiers
 1..651
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:5357472"
 /clone_1ib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 154 a 198 c 149 g 150 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 651;
 Best Local Similarity 96.0%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGGCTGAGAGCTCTTCCA 25
 |||||||
 Db 61 GCTTCTTGGCTGAGAGCTCTTCCA 85

RESULT 7
 B1730180 656 bp mRNA linear EST 20-SEP-2001
 LOCUS 603349724F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5357429 5',
 DEFINITION mRNA sequence.

ACCESSION B1730180
 VERSION B1730180.1 GI:15707193
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 656)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM11908 row: p column: 06
 High quality sequence stop: 649.

FEATURES
 source 1..656
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_image="5357429"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 142 a 210 c 152 g 152 t
 ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 656;
 Best Local Similarity 96.0%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGGCTGAGAGCTCTTCCA 25
 |||||||
 Db 402 GCTTCTTGGCTGAGAGCTCTTCCA 426

RESULT 8
 B1729678 663 bp mRNA linear EST 20-SEP-2001
 LOCUS 603349362F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5356959 5',
 DEFINITION mRNA sequence.

ACCESSION B1729678
 VERSION B1729678.1 GI:15706691
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 663)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM11907 row: 1 column: 16
 High quality sequence stop: 659.

FEATURES
 source 1..663
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_image="5356959"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 156 a 209 c 151 g 147 t
 ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 663;
 Best Local Similarity 96.0%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTGGCTGAGAGCTCTTCCA 25
 |||||||
 Db 152 GCTTCTTGGCTGAGAGCTCTTCCA 176

RESULT 9
 BG298553 672 bp mRNA linear EST 21-FEB-2001
 LOCUS 602336937F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4511697 5',
 DEFINITION mRNA sequence.

ACCESSION BG298553
 VERSION BG298553.1 GI:13063322
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 672)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10395 row: 1 column: 10
 High quality sequence stop: 666.

FEATURES
 source 1..672
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"

/clone="IMAGE:4511697"
/clone_lib="NIH_MGC_94"
/issue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: PCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 150 a 206 c 156 g 160 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 12; Length 672;
Best Local Similarity 96.0%; Pred. No. 24;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCTGAGAGCTCTTCCA 25
|||||
DB 341 GCTTCTTTGCTGAGAGCTCTTCCA 365

RESULT 10
BI736996 676 bp mRNA linear EST 20-SEP-2001
LOCUS 603360842F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5367875 5',
DEFINITION mRNA sequence.
ACCESSION BI736996
VERSION BI736996.1 GI:15714009
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 676)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM1936 row: c column: 12
High quality sequence stop: 662.

FEATURES

Location/Qualifiers
1..676
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5367875"
/clone_lib="NIH_MGC_94"
/issue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: PCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT

144 a 214 c 160 g 158 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 676;
Best Local Similarity 96.0%; Pred. No. 24;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCTGAGAGCTCTTCCA 25
|||||
DB 409 GCTTCTTTGCTGAGAGCTCTTCCA 433

RESULT 11
BI735044 681 bp mRNA linear EST 20-SEP-2001
LOCUS 603356176F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5363395 5',
DEFINITION mRNA sequence.
ACCESSION BI735044
VERSION BI735044.1 GI:15712057
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 681)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM1924 row: h column: 20
High quality sequence stop: 680.

FEATURES

Location/Qualifiers
1..681
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5363395"
/clone_lib="NIH_MGC_94"
/issue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: PCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dt primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 152 a 220 c 156 g 152 t 1 others

Query Match 93.6%; Score 23.4; DB 13; Length 681;
Best Local Similarity 96.0%; Pred. No. 24;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTTCTTTGCTGAGAGCTCTTCCA 25
|||||
DB 244 GCTTCTTTGCTGAGAGCTCTTCCA 268

RESULT 12

BI295809 682 bp mRNA linear EST 21-FEB-2001
LOCUS 602939229F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4505134 5',
DEFINITION mRNA sequence.
ACCESSION BI295809
VERSION BI295809.1 GI:13057815
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

ORIGIN

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 682)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Place: LLAM10378 row: g column: 23
 High quality sequence stop: 603.
 Location/Qualifiers

FEATURES

source

1..682
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4505134"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 137 a 205 c 167 g 173 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 12; Length 682;
 Best Local Similarity 96.0%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCTTCTTTGCTGAGAGCTCTTCCA 25

Db 512 GCTTCTTTGCTGAGAGCTCTTCCA 536

RESULT 13

BI737008

DEFINITION BI737008 694 bp mRNA linear EST 20-SEP-2001
 60336085F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5368117 5',
 mRNA sequence.

ACCESSION BI737008

VERSION BI737008.1 GI:15714021

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
 1 (bases 1 to 694)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE Unpublished (1999)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM11936 row: m column: 14
 High quality sequence stop: 694.
 Location/Qualifiers

FEATURES

source

1..694
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:5368117"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 157 a 216 c 156 g 165 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 694;
 Best Local Similarity 96.0%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCTTCTTTGCTGAGAGCTCTTCCA 25

Db 363 GCTTCTTTGCTGAGAGCTCTTCCA 387

RESULT 14

BG298167

DEFINITION BG298167 696 bp mRNA linear EST 21-FEB-2001
 60239630F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4507824 5',
 mRNA sequence.

ACCESSION BG298167

VERSION BG298167.1 GI:13062367

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
 1 (bases 1 to 696)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE Unpublished (1999)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM10385 row: h column: 01
 High quality sequence stop: 677.
 Location/Qualifiers

FEATURES

source

1..696
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4507824"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 148 a 218 c 164 g 166 t

ORIGIN

Query Match 93.6%; Score 23.4; DB 12; Length 696;
 Best Local Similarity 96.0%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCTTCTTTGCTGAGAGCTCTTCCA 25

Db 459 GCTTCTTTGCTGAGAGCTCTTCCA 483

RESULT 15

BI735675

DEFINITION BI735675 697 bp mRNA linear EST 20-SEP-2001
 60335782F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5364964 5',
 mRNA sequence.

ACCESSION BI735675

VERSION BI735675.1 GI:15712688

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 697)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLAM1928 row: 1 column: 05
High quality sequence stop: 697.

FEATURES
SOURCE

Location/Qualifiers
1..697
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5364964"
/clone_id="NIH-MGC.94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT 136 a 209 c 168 g 184 t
ORIGIN

Query Match 93.6%; Score 23.4; DB 13; Length 697;
Best Local Similarity 96.0%; Pred. No. 24;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GCTTCTTGCTGAGACTCTCA 25
|||||
Db 585 GCTTCTTGCTGAGACTCTCA 609

Search completed: March 17, 2003, 13:09:23
Job time : 886.737 secs

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 09:43:42 ; Search time 202.495 Seconds

(Without alignments)
3161.870 Million cell updates/sec

Title: US-09-836-439-6

Perfect score: 22

Sequence: 1 aagactctgagtaacaa 22

Scoring table:

IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_hlg: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hlg_hum: *
31: em_hlg_inv: *
32: em_hlg_other: *
33: em_hlg_mus: *
34: em_hlg_pln: *
35: em_hlg_rtd: *
36: em_hlg_mam: *
37: em_hlg_vrt: *
38: em_mv: *
39: em_hlgo_hum: *
40: em_hlgo_mus: *
41: em_hlgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	88.2	17938	AF198437	AF198437 Theileria
2	18.8	85.5	42208	AL589989	AL589989 Human DNA
3	18.8	85.5	116685	HS516C23	293021 Human DNA s
4	18.8	85.5	178168	AC012350	AC012350 Homo sapi
5	18.8	85.5	180970	AC009302	AC009302 Homo sapi
6	18.8	85.5	182152	AC067929	AC067929 Homo sapi
7	18.8	85.5	189281	AC012458	AC012458 Homo sapi
8	18.8	85.5	191434	AC087477	AC087477 Homo sapi
9	18.8	85.5	197437	AL772303	AL772303 Homo sapi
10	18.8	85.5	223761	AC127239	AC127239 Mus muscu
11	18.8	85.5	229247	AC098888	AC098888 Mus muscu
12	18.8	85.5	317515	AC099415	AC099415 Mus muscu
13	18.4	83.6	10243	AE011043	AE011043 Methanosa
14	18.4	83.6	10467	AE010860	AE010860 Methanosa
15	18.4	83.6	152556	CNS01DSY	AL122035 Human chr
16	18.4	83.6	153951	AC019027	AC019027 Homo sapi
17	18.4	81.8	148507	AC011263	AC011263 Homo sapi
18	18.4	81.8	158633	AC015472	AC015472 Homo sapi
19	18.4	81.8	178199	AC018772	AC018772 Homo sapi
20	18.4	81.8	224068	CNS01DUU	AL133246 BAC sequ
21	17.8	80.9	907	AF213436	AF213436 Homo sapi
22	17.8	80.9	927	107695	107695 Sequence 18
23	17.8	80.9	3159	HS05SE1	X51934 Human dyslr
24	17.8	80.9	57307	AL603652	AL603652 Human DNA
25	17.8	80.9	83007	AC097448	AC097448 Homo sapi
26	17.8	80.9	107406	AC013491	AC013491 Homo sapi
27	17.8	80.9	141984	F911	AC007591 Arabidops
28	17.8	80.9	148571	AC105316	AC105316 Homo sapi
29	17.8	80.9	161874	AC079864	AC079864 Homo sapi
30	17.8	80.9	163132	AC097180	AC097180 Rattus no
31	17.8	80.9	176287	AL513347	AL513347 Mouse DNA
32	17.8	80.9	179591	AC093902	AC093902 Homo sapi
33	17.8	80.9	182463	AC120773	AC120773 Rattus no
34	17.8	80.9	182933	AC097610	AC097610 Rattus no
35	17.8	80.9	184573	AC111144	AC111144 Mus muscu
36	17.8	80.9	191664	AC125751	AC125751 Rattus no
37	17.8	80.9	200559	AC1121805	AC1121805 Mus muscu
38	17.8	80.9	201296	AC115299	AC115299 Mus muscu
39	17.8	80.9	203753	AC104343	AC104343 Homo sapi
40	17.8	80.9	207945	AC117841	AC117841 Rattus no
41	17.8	80.9	207974	ALB31771	ALB31771 Mus muscu
42	17.8	80.9	208953	CNS01MR3	AL160314 Human chr
43	17.8	80.9	210779	AC007450	AC007450 Homo sapi
44	17.4	79.1	12024	AE000995	AE000995 Archaeogl
45	17.4	79.1	65499	AC100433	AC100433 Mus muscu

ALIGNMENTS

RESULT 1
AF198437
LOCUS
DEFINITION
AF198437 17938 bp DNA linear INV 05-DEC-2000
Theileria parva strain Muguga hypothetical telomeric SfiI fragment
20 protein 3, hypothetical telomeric SfiI fragment 20 protein 2,
and hypothetical telomeric SfiI fragment 20 protein 1 genes,
complete cds.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AF198437 L36964
AF198437.1 GI:11545211
Theileria parva.
Theileria parva.
Alveolata; Apicomplexa; Piroplasmida; Theileridae;
Theileria.
1 (bases 17114 to 17938)

repeat_L1region	/note="L1MC1 repeat: matches 5398. .6314 of consensus 23343. .23754
repeat_L1region	/note="L2 repeat: matches 2029. .2530 of consensus" 24063. .24264
repeat_L1region	/note="L1B repeat: matches 189. .390 of consensus" 24302. .24445
repeat_L1region	/note="22 copies 2 mer tt 75% conserved" 24361. .30507
repeat_L1region	/note="L1P2 repeat: matches 4. .6144 of consensus" 30510. .30679
repeat_L1region	/note="L1B1 repeat: matches 1. .170 of consensus" 31272. .31782
repeat_L1region	/note="L2 repeat: matches 960. .2750 of consensus" 31811. .32018
repeat_L1region	/note="MIR repeat: matches 39. .262 of consensus" 32300. .33151
repeat_L1region	/note="L1P1 repeat: matches 5333. .6155 of consensus" 33152. .33536
repeat_L1region	/note="L1NSP4 repeat: matches 192. .276 of consensus" 33484. .33868
repeat_L1region	/note="L2 repeat: matches 1862. .2267 of consensus" 33887. .34010
repeat_L1region	/note="L1R3 repeat: matches 1. .143 of consensus" 34187. .34218
repeat_L1region	/note="16 copies 2 mer aa 84% conserved" 34220. .34364
repeat_L1region	/note="L1R3 repeat: matches 382. .517 of consensus" 34403. .34964
repeat_L1region	/note="L1MEC repeat: matches 272. .906 of consensus" 35050. .35158
repeat_L1region	/note="L1M4 repeat: matches 6176. .6290 of consensus" 35141. .35266
repeat_L1region	/note="L1M4 repeat: matches 2121. .2241 of consensus" 35267. .35587
repeat_L1region	/note="A1U1 repeat: matches 1. .309 of consensus" 35588. .36308
repeat_L1region	/note="L1M4 repeat: matches 2241. .2996 of consensus" 36248. .36925
repeat_L1region	/note="L1ME2 repeat: matches 5476. .6154 of consensus" 36950. .37121
repeat_L1region	/note="L1MEC repeat: matches 1782. .1975 of consensus" 37125. .37741
repeat_L1region	/note="L1M4 repeat: matches 2359. .3033 of consensus" 37749. .37877
repeat_L1region	/note="A1USq/x repeat: matches 1. .129 of consensus" 37878. .37923
repeat_L1region	/note="23 copies 2 mer ca 100% conserved" 37928. .38256
repeat_L1region	/note="A1USx repeat: matches 1. .300 of consensus" 38262. .38433
repeat_L1region	/note="L1M4 repeat: matches 3047. .3335 of consensus" 38463. .38837
repeat_L1region	/note="L1M4 repeat: matches 3790. .4186 of consensus" 38847. .39138
repeat_L1region	/note="A1U0 repeat: matches 12. .293 of consensus" 39334. .39389
repeat_L1region	/note="23 copies 2 mer aa 76% conserved" 40977. .41071
repeat_L1region	/note="MIR repeat: matches 12. .120 of consensus" 41529. .41819
repeat_L1region	/note="L2 repeat: matches 1780. .2071 of consensus" 41800. .41805
misc_feature	/note="153 transposable element excised from this position" 41820. .42230
repeat_L1region	/note="M1U1 repeat: matches 56. .505 of consensus" 42267. .42627
repeat_L1region	/note="L2 repeat: matches 2301. .2677 of consensus" 42644. .42731
repeat_L1region	/note="44 copies 2 mer aa 69% conserved" 43499. .43640
repeat_L1region	/note="71 copies 2 mer tt 63% conserved" 44092. .44158

```

repeat_region /note="LTR1D repeat: matches 1. .66 of consensus"
44166. .44649
/note="LTR1D repeat: matches 5. .505 of consensus"
repeat_region 44686. .45053
/note="LTR1D repeat: matches 113. .503 of consensus"
repeat_region 45793. .45960
/note="MIR repeat: matches 7. .171 of consensus"
repeat_region 46079. .46271
/note="LTR1F repeat: matches 347. .541 of consensus"
repeat_region 46316. .46363
/note="LTR2CB repeat: matches 236. .283 of consensus"
repeat_region 46602. .46641
/note="20 copies 2 mer tg 95% conserved"
misc_feature complement(47028. .47443)
/note="match: GSS B53031 clone 2008G6"
repeat_region 47509. .48290
/note="L1MB5 repeat: matches 535. .6166 of consensus"
repeat_region 48283. .49518
/note="L1MB5 repeat: matches 3964. .5271 of consensus"
repeat_region 49519. .49586
/note="34 copies 2 mer ta 72% conserved"
repeat_region 49980. .50288
/note="AluSq repeat: matches 3. .311 of consensus"
repeat_region 50661. .51118
/note="L1MC4 repeat: matches 6102. .7977 of consensus"
misc_feature complement(51869. .52243)
repeat_region /note="match: GSS AQ109033 clone 2377A15"
52215. .52242
/note="14 copies 2 mer ac 100% conserved"
misc_feature <52243. .>52703
repeat_region /note="match: GSS AQ234004"
54296. .54852
/note="12 repeat: matches 2180. .2743 of consensus"
repeat_region 55942. .55097
/note="MIR repeat: matches 49. .210 of consensus"

```

Query Match	85.5%	Score 18.8;	DB 9;	Length 116685;
Best Local Similarity	90.9%	Pred. NO. 90;		
Matches 20; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

```

QY      1  AAGACTTCTGAGTAACAATCAA  22
          || ||||| ||||| |||||
Db 59346 AACTCTTCTGAGTAACAATCAA  59325

```

LOCUS	178168 bp	DNA	linear	HTG 01-APR-2000
DEFINITION	Homo sapiens clone RP11-16N9, WORKING DRAFT SEQUENCE, 16 unordered pieces.			
ACCESSION	AC012350			
VERSION	AC012350.3	GI:7381803		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (bases 1 to 178168)	Bliren,B., Linton,L., Nusbaum,C. and Lander,E.	Homo sapiens, clone RP11-16N9	Unpublished	2 (bases 1 to 178168)
Bliren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,				

Batten, B., Lincoln, L., Neudbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barua, N., Beckerly, R., Boguslavsky, L., Bouknight, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArillano, K., Dewar, K., Domingo, M., Donelan, D., Doyle, M., Ferrara, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gallegan, J., Gadyana, S., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lohoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, D., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

TITLE Direct Submission
JOURNAL Submitted (13-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 180970)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 180970)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 180970)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 28, 2000 this sequence version replaced gi:5732171.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0071J24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-350124: the clone sequenced to the right is RP11-296A19. Actual start of this clone is at base position 1 of RP11-71J24; actual end is at base position 180970 of RP11-71J24.

FEATURES

source

Location/Qualifiers
1..180970
/Organism="Homo sapiens"
/db_xref="taxon:9606"
/Chromosome="2"
/map="2"
/clone="RP11-71J24"
/clone_1fb="RPCI-11"

repeat_region	4266..4361	/rpt_family="MIR"
repeat_region	7373..7538	/rpt_family="MIR"
repeat_region	7878..8175	/rpt_family="Alu"
repeat_region	8215..8281	/rpt_family="MIR"
repeat_region	8699..8778	/rpt_family="MIR"
repeat_region	9014..9113	/rpt_family="L1"
repeat_region	9120..9530	/rpt_family="L1"
repeat_region	9639..9748	/rpt_family="L2"
repeat_region	11531..11729	/rpt_family="MIR"
repeat_region	11735..11888	/rpt_family="L1"
repeat_region	12416..12459	/rpt_family="L2"
repeat_region	12608..12966	/rpt_family="MIR"
repeat_region	14055..14365	/rpt_family="Alu"
repeat_region	14659..14822	/rpt_family="MIR"
repeat_region	15167..15466	/rpt_family="Alu"
repeat_region	15865..15942	/rpt_family="L2"
repeat_region	16226..16552	/rpt_family="Alu"
repeat_region	17132..17438	/rpt_family="Alu"
repeat_region	19203..19508	/rpt_family="Alu"
repeat_region	19511..19818	/rpt_family="Alu"
repeat_region	20445..20594	/rpt_family="Alu"
repeat_region	20732..21129	/rpt_family="ERV1"
repeat_region	22146..22254	/rpt_family="MER1_type"
repeat_region	23476..23592	/rpt_family="MIR"
repeat_region	24585..24914	/rpt_family="L1"
repeat_region	25584..25744	/rpt_family="MIR"
repeat_region	25798..26320	/rpt_family="MER1_type"
repeat_region	29905..30145	/rpt_family="MIR"
repeat_region	30933..31063	/rpt_family="ERV1"
repeat_region	31064..31118	/rpt_family="ERV1"
repeat_region	31135..31192	/rpt_family="ERV1"
repeat_region	31213..31623	/rpt_family="ERV1"
repeat_region	31690..32052	/rpt_family="MIR"
repeat_region	32252..32408	/rpt_family="MIR"
repeat_region	32638..32993	/rpt_family="MIR"
repeat_region	33136..33707	/rpt_family="MIR2_type"
repeat_region	35611..35898	

2 (bases 1 to 182152) Nusbaum, C., Lander, E., Abraham, H., Allen, N., Birren, B., Linton, L., Baldwin, J., Barna, N., Bastien, V., Bedalov, J., Anderson, S., Baldwin, J., Bonkshgalter, B., Brown, A., Burkett, G., Boguski, M., Boudreau, A., Castle, A., Chepel, Y., Colangelo, M., Collins, S., Campolongo, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Gage, D., Collymore, A., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Dode, S., Domino, M., Ginde, S., Goyette, M., Graham, L., Galagan, J., Gardyna, S., Grant, G., Haeos, B., Heaford, A., Horton, L., Grand-Pierre, N., Grant, G., Haeos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lebecky, J., Klein, J., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., Levine, R., McEwan, P., McGuck, A., McKernan, K., McPeckers, R., McElm, J., Menus, F., Milnova, T., Miranda, C., Mlenka, V., Morrow, J.,

*	1	2222:	contig of 2222 bp in length
*	2223	2322:	gap of 100 bp
*	2232	3785:	contig of 1463 bp in length
*	3786	3885:	gap of 100 bp
*	3886	6753:	contig of 2868 bp in length

[illegible]

```

*----- Genome Center -----*
* Center: Washington University Genome Sequencing Center *
* Center code: WUGSC *
* Web site: http://genome.wustl.edu/gsc/index.shtml *
*----- Project Information -----*
* Center project name: H.NH0522b15 *
*----- Summary Statistics -----*
* Sequencing vector: M13; 77% *
* Sequencing vector: plasmid; 21% *
* Chemistry: Dye-primer ET; 77% of reads *
* Chemistry: Dye-terminator Big Dye; 23% of reads *
* Assembly program: Phrap; version 0.990319 *
* Consensus quality: 183385 bases at least Q40 *
* Consensus quality: 185502 bases at least Q30 *
* Consensus quality: 186573 bases at least Q20 *
* Insert size: 197000; agarose-fp *
* Insert size: 188381; sum-of-ctrls *
* Quality coverage: 4.50 in Q20 bases; agarose-fp *
* Quality coverage: 4.71 in Q20 bases; sum-of-ctrls *
*-----*

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1      4603: contig of 4603 bp in length
*      4604      4703: gap of unknown length
*      4704      9861: contig of 5038 bp in length
*      9762      9861: gap of unknown length

```

```

*      *      *      *      *
*      9862      18746: contig of 8885 bp in length
*      18747      18846: gap of unknown length
*      18847      27396: contig of 8550 bp in length
*      27397      27496: gap of unknown length
*      27497      35489: contig of 7993 bp in length
*      35490      35589: gap of unknown length
*      35590      51472: contig of 15883 bp in length
*      51473      51572: gap of unknown length
*      51573      65229: contig of 13657 bp in length
*      65230      65329: gap of unknown length
*      65330      89666: contig of 24337 bp in length
*      89667      89766: gap of unknown length
*      89767      117943: contig of 28177 bp in length
*      117944      118044: gap of unknown length
*      118044      189281: contig of 71238 bp in length.
*      Location/Qualifiers
*      source
*      1. 189281
*      /organism="Homo sapiens"
*      /db_xref="taxon:9606"
*      /chromosome="15"
*      /clone="RP11-522B15"
*      misc_feature
*      1. 4603
*      /note="assembly_name:Contig6"
*      misc_feature
*      4704. 9761
*      /note="assembly_name:Contig7"
*      misc_feature
*      9862. 18746
*      /note="assembly_name:Contig8"
*      misc_feature
*      18847. 27396
*      /note="assembly_name:Contig9"
*      misc_feature
*      27497. 35489
*      /note="assembly_name:Contig10"
*      misc_feature
*      35590. 51472
*      /note="assembly_name:Contig11"
*      misc_feature
*      51573. 65229
*      /note="assembly_name:Contig12"
*      misc_feature
*      clone_end:Sp6
*      vector_side:right"
*      misc_feature
*      65330. 89666
*      /note="assembly_name:Contig13"
*      misc_feature
*      89767. 117943
*      /note="assembly_name:Contig14"
*      misc_feature
*      clone_end:T7
*      vector_side:left"
*      misc_feature
*      118044. 189281
*      /note="assembly_name:Contig15"
*      misc_feature
*      54120 a 39374 c 39106 g 55769 t 912 others
*      BASE COUNT
*      ORIGIN

```

```

Query Match      85.5%; Score 18.8; DB 2; Length 189281;
Best Local Similarity 90.9%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY      1 AAGACTTCTGAGTACATCA 22
Db 181216 AAGACTTCTGATTAACATGAA 181237
RESULT 8
AC087477      191434 bp      DNA      linear      PRI 16-JUL-2002
LOCUS      Homo sapiens chromosome 15, clone RP11-522B15, complete sequence.
AC087477
AC087477.8      GI:21844630
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 191434)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 15, clone RP11-522B15
JOURNAL
REFERENCE
2 (bases 1 to 191434)

```

AUTHORS
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
 Camarata,J., Campoliano,A., Choapel,Y., Colangelo,M., Collins,S.,
 Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
 Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Karatas,A., Larocque,K., Lamazares,R., Landers,T.,
 Lenockzy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
 Matthews,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
 McNeus,L., Mihova,T., Mlenga,V., Murphy,T., Nguyen,C.,
 Mcheters,R., Meldrim,J., Menes,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
 Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
 Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,
 Soungaz,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strausz,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE
 Direct Submission
 JOURNAL
 Submitted (03-JAN-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE
 3 (bases 1 to 191434)

AUTHORS
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Boukhgalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
 Chazaro,B., Choapel,Y., Colangelo,M., Collins,S., Collamore,A.,
 Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., Fitzgerld,M., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
 Lamazares,R., Landers,T., Lenockzy,J., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
 Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
 McNeus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
 Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
 Schnupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Strausz,N., Subramanian,A., Talamas,J., Testaye,S.,
 Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
 Direct Submission
 JOURNAL
 Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 REFERENCE
 4 (bases 1 to 191434)

AUTHORS
 Birren,B., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
 Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choapel,Y., Collamore,A.,
 Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., Fitzgerld,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Menes,L., Mihova,T., Mlenga,V.,
 McCarthy,M., Meldrim,J., Menes,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE
 Direct Submission
 JOURNAL
 Submitted (16-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT
 On Jul 16, 2002 this sequence version replaced gi:20043143.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/BM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11935

Center clone name: 522_B_15

FEATURES

```

source
    Location/Qualifiers
    1..191434
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="15"
    /map="15"
    /clone="RP11-522B15"
    /clone_id="RRC1-11 Human Male BAC"
    1..777
    /note="PCR product sequence only"
repeat_region
    2518..2556
    /rpt_family="MIR"
repeat_region
    3963..4011
    /rpt_family="(CA)n"
repeat_region
    4103..4186
    /rpt_family="C-rich"
repeat_region
    4898..4965
    /rpt_family="CT-rich"
repeat_region
    5152..5213
    /rpt_family="G-rich"
repeat_region
    5215..5241
    /rpt_family="G-rich"
repeat_region
    6304..6308
    /rpt_family="(GA)n"
unsure
    /note="<30 qual SINGL region"
    6363..6569
    /note="<30 qual SINGL region"
    7379..7470
    /rpt_family="CT-rich"
repeat_region
    7515..7548
    /rpt_family="(TG)n"
repeat_region
    7549..7602
    /rpt_family="(GA)n"
repeat_region
    10343..10363
    /rpt_family="(TGGGGG)n"
unsure
    /note="<30 qual SINGL region"
    10422..10428
    /note="<30 qual SINGL region"
    11032..11060
    /rpt_family="GC-rich"
repeat_region
    11925..11954
    /rpt_family="(T)n"
repeat_region
    13905..13926
    /rpt_family="(TC)n"
unsure
    /note="<30 qual SINGL region"
    complement(14459..14464)
unsure
    /note="<30 qual SINGL region"
    complement(14475..14479)
unsure
    /note="<30 qual SINGL region"
    complement(14493..14498)
    /note="<30 qual SINGL region"
    18818..19134
    /rpt_family="AluSq"
repeat_region
    complement(20563..20758)
    /rpt_family="MIR"
repeat_region
    complement(20896..21016)
    /rpt_family="MIR"
repeat_region
    21079..21098
    /rpt_family="(CA)n"
repeat_region
    21150..21200
    /rpt_family="(TATANG)n"
repeat_region
    complement(21400..21579)
    /rpt_family="MIR"
repeat_region
    21586..21876

```

```

repeat_region
    /rpt_family="AluSx"
    complement(22988..23113)
repeat_region
    /rpt_family="L2"
    complement(23359..23666)
repeat_region
    /rpt_family="AluSq"
    complement(23784..24062)
repeat_region
    /rpt_family="AluUo"
    24327..24350
repeat_region
    /rpt_family="(A)n"
    24471..24597
repeat_region
    /rpt_family="L2"
    25115..25363
repeat_region
    /rpt_family="AluUb"
    complement(25426..25712)
repeat_region
    /rpt_family="AluSc"
    26549..26667
repeat_region
    /rpt_family="FLAM_A"
    complement(26714..27135)
repeat_region
    /rpt_family="AluUo"
    27835..27885
repeat_region
    /rpt_family="AT-rich"
    complement(29173..29331)
repeat_region
    /rpt_family="AluUo"
    complement(29332..30777)
repeat_region
    /rpt_family="L1PAS"
    complement(30778..30790)
repeat_region
    /rpt_family="AluUo"
    complement(30791..31103)
repeat_region
    /rpt_family="AluSp"
    complement(31104..31235)
repeat_region
    /rpt_family="AluUo"
    31242..31310
repeat_region
    /rpt_family="MLT1P1"
    31437..31741
repeat_region
    /rpt_family="MLT1P1"
    31760..31821
repeat_region
    /rpt_family="L2"
    31829..31851
repeat_region
    /rpt_family="AT-rich"
    complement(31951..31993)

```

Query Match 85.5%; Score 18.8; DB 9; Length 191434;

Best local similarity 90.9%; Pred. No. 85;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGACTTCTGACTACATCA 22

DB 137915 AAGACTTCTGATTACATCA 137936

RESULT 9

AL772303/c 197437 bp DNA linear HTG 13-JUL-2002

LOCUS Mus musculus chromosome 2 clone RP23-185P20, *** SEQUENCING IN

DEFINITION AL772303

ACCESSION AL772303 3 GI:21748303

VERSION HTG: HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTROP.

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 197437)

AUTHORS Almeida, J.

TITLE Submitted (12-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,

JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Jul 14, 2002 this sequence version replaced g1:21540206.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>


```

misc_feature      /note="assembly_name:Contig12"
                  48422..118533
misc_feature      /note="assembly_name:Contig13"
                  118634..212339
misc_feature      /note="assembly_name:Contig14"
                  212440..216233
misc_feature      /note="assembly_name:Contig18"
                  216334..223761
misc_feature      /note="assembly_name:Contig19"
                  71420 a 41598 c 40871 g 69263 t 609 others
BASE COUNT
ORIGIN

```

```

Query Match      85.5%: Score 18.8; DB 2; Length 223761;
Best Local Similarity 90.9%: Pred. No. 84;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 AAGACTTCTGAGTACATCA 22
    |||||  |||  |||||
Db 87767 AAGACTTCTGAGTACATCA 87788

```

```

RESULT 11
AC098888      229247 bp  DNA  linear  ROD 21-JUN-2002
DEFINITION    Mus musculus clone RP23-122N8, complete sequence.
ACCESSION     AC098888
VERSION       AC098888.4  GI:19909498
KEYWORDS      HTG.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS       McPherson,J.D. and Waterston,R.H.
TITLE         The sequence of Mus musculus clone
JOURNAL       Unpublished

```

```

REFERENCE
AUTHORS       2 (bases 1 to 229247)
TITLE         McPherson,J.D. and Waterston,R.H.
JOURNAL       Direct Submission
AUTHORS       Submitted (05-NOV-2001) Genome Sequencing Center, 4444 Forest Park
TITLE         Parkway, St. Louis, MO 63108, USA
JOURNAL       3 (bases 1 to 229247)

```

```

REFERENCE
AUTHORS       McPherson,J.D. and Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (03-APR-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS       Parkway, St. Louis, MO 63108, USA
TITLE         4 (bases 1 to 229247)
JOURNAL       McPherson,J.D. and Waterston,R.H.
AUTHORS       Direct Submission
TITLE         Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL       Parkway, St. Louis, MO 63108, USA

```

```

COMMENT
On Apr 3, 2002 this sequence version replaced gi:17105321.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0122N08
-----
FEATURES
source
1..229247
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="RP23-122N8"
BASE COUNT  70714 a 42415 c 42708 g 73410 t
ORIGIN

```

```

Query Match      85.5%: Score 18.8; DB 10; Length 229247;
Best Local Similarity 90.9%: Pred. No. 83;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 AAGACTTCTGAGTACATCA 22
    |||||  |||  |||||
Db 83202 AAGACTTCTGAGTACATCA 83223

```

```

RESULT 12
AC099415      317515 bp  DNA  linear  HTG 05-JUN-2002
DEFINITION    Mus musculus chromosome UNK clone RP23-122D8, WORKING DRAFT
SEQUENCE, 56 unordered pieces.
ACCESSION     AC099415
VERSION       AC099415.3  GI:21326409
KEYWORDS      HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS       1 (bases 1 to 317515)
TITLE         McPherson,J.D. and Waterston,R.H.
JOURNAL       The sequence of Mus musculus clone

```

```

REFERENCE
AUTHORS       2 (bases 1 to 317515)
TITLE         McPherson,J.D. and Waterston,R.H.
JOURNAL       Unpublished
AUTHORS       2 (bases 1 to 317515)
TITLE         McPherson,J.D. and Waterston,R.H.
JOURNAL       Direct Submission
AUTHORS       Submitted (14-NOV-2001) Genome Sequencing Center, 4444 Forest Park
TITLE         Parkway, St. Louis, MO 63108, USA
JOURNAL       3 (bases 1 to 317515)
AUTHORS       McPherson,J.D. and Waterston,R.H.
TITLE         Direct Submission
JOURNAL       Submitted (05-JUN-2002) Genome Sequencing Center, 4444 Forest Park
AUTHORS       Parkway, St. Louis, MO 63108, USA
COMMENT
On Jun 5, 2002 this sequence version replaced gi:20069750.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0122D08
-----
----- Summary Statistics -----
Sequencing vector: M13; 168
Sequencing vector: plasmid; 848
Chemistry: Dye-primer; 0% of reads
Chemistry: Dye-terminator; Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 293243 bases at least Q40
Consensus quality: 303369 bases at least Q30
Consensus quality: 309642 bases at least Q20
Insert size: 224000; agarose-fp
Insert size: 322378; sum-of-ctrls
Quality coverage: 37.21 in Q20 bases; agarose-fp
Quality coverage: 24.94 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1211: contig of 1211 bp in length
* 1212 1311: gap of unknown length
* 1312 2357: contig of 1046 bp in length
* 2358 2457: gap of unknown length
* 2458 3574: contig of 1117 bp in length
* 3574 3674: gap of unknown length
* 3674 4817: contig of 1143 bp in length
* 4817 4917: gap of unknown length
* 4917 6015: contig of 1098 bp in length
* 6015 6016: gap of unknown length

```

```

* 6116 7259: contig of 1144 bp in length
* 7260 7359: gap of unknown length
* 7360 8831: contig of 1472 bp in length
* 8832 8931: gap of unknown length
* 8932 10179: contig of 1248 bp in length
* 10180 10279: gap of unknown length
* 10280 11586: contig of 1307 bp in length
* 11587 11686: gap of unknown length
* 11687 12940: contig of 1254 bp in length
* 12941 13040: gap of unknown length
* 13041 14154: contig of 1114 bp in length
* 14155 14254: gap of unknown length
* 14255 15728: contig of 1474 bp in length
* 15729 15828: gap of unknown length
* 15829 17212: contig of 1384 bp in length
* 17213 17312: gap of unknown length
* 17313 18620: contig of 1308 bp in length
* 18621 18720: gap of unknown length
* 18721 20283: contig of 1563 bp in length
* 20284 20384: gap of unknown length
* 20384 21676: contig of 1293 bp in length
* 21677 21776: gap of unknown length
* 21777 23147: contig of 1371 bp in length
* 23148 23247: gap of unknown length
* 23248 24781: contig of 1534 bp in length
* 24782 24881: gap of unknown length
* 24882 26627: contig of 1746 bp in length
* 26628 26727: gap of unknown length
* 26728 28451: contig of 1724 bp in length
* 28452 28551: gap of unknown length
* 28552 29863: contig of 1312 bp in length
* 29864 29963: gap of unknown length
* 29964 31941: contig of 1978 bp in length
* 31942 32041: gap of unknown length
* 32042 33660: contig of 1619 bp in length
* 33661 33760: gap of unknown length
* 33761 34889: contig of 1129 bp in length
* 34890 36451: gap of unknown length
* 36452 36551: gap of unknown length
* 36552 38157: contig of 1606 bp in length
* 38158 38257: gap of unknown length
* 38258 40052: contig of 1795 bp in length
* 40053 40152: gap of unknown length
* 40153 41725: contig of 1573 bp in length
* 41726 41825: gap of unknown length
* 41826 42986: contig of 1171 bp in length
* 42987 43096: gap of unknown length
* 43097 44639: contig of 1543 bp in length
* 44640 44739: gap of unknown length
* 44740 46047: contig of 1308 bp in length
* 46048 46147: gap of unknown length
* 46148 47831: contig of 1684 bp in length
* 47832 47931: gap of unknown length
* 47932 49690: contig of 1759 bp in length
* 49691 49790: gap of unknown length
* 49791 51094: contig of 1304 bp in length
* 51095 51194: gap of unknown length
* 51195 53501: contig of 2307 bp in length
* 53502 53601: gap of unknown length
* 53602 55272: contig of 1671 bp in length
* 55273 55372: gap of unknown length
* 55373 57428: contig of 2056 bp in length
* 57429 59819: contig of 2291 bp in length
* 59820 59919: gap of unknown length
* 59920 61717: contig of 1798 bp in length
* 61718 61817: gap of unknown length
* 61818 64311: contig of 2494 bp in length
* 64312 67042: contig of 2631 bp in length
* 67043 67142: gap of unknown length
* 67143 69238: contig of 2096 bp in length

```

```

* 69239 69338: gap of unknown length
* 69339 71520: contig of 2182 bp in length
* 71521 71620: gap of unknown length
* 71621 74319: contig of 2699 bp in length
* 74320 74419: gap of unknown length
* 74420 76137: contig of 1718 bp in length
* 76138 76237: gap of unknown length
* 76238 79611: contig of 3374 bp in length
* 79612 82702: contig of 2991 bp in length
* 82703 82802: gap of unknown length
* 82803 84448: contig of 1646 bp in length
* 84449 84548: gap of unknown length
* 84549 87522: contig of 2874 bp in length
* 87523 89088: contig of 1566 bp in length
* 89089 89188: gap of unknown length
* 89189 92189: contig of 3001 bp in length
* 92190 92289: gap of unknown length
* 92290 96604: contig of 4315 bp in length
* 96605 96705: gap of unknown length
* 96706 98834: contig of 2130 bp in length
* 98835 98934: gap of unknown length
* 98935 109620: contig of 10686 bp in length
* 109621 109720: gap of unknown length
* 109721 177250: contig of 67530 bp in length
* 177251 177350: gap of unknown length
* 177351 317515: contig of 140165 bp in length.

```

FEATURES

source

```

misc_feature      1..1211
                  /organism="Mus musculus"
                  /db_xref="taxon:10090"
                  /chromosome="X"
                  /clone="RP23-122D8"
misc_feature      1312..2357
                  /note="assembly_name:Contig189"
misc_feature      2458..3574
                  /note="assembly_name:Contig193"
misc_feature      3675..4817
                  /note="assembly_name:Contig141"
misc_feature      4916..6015
                  /note="assembly_name:Contig142"
misc_feature      6116..7259
                  /note="assembly_name:Contig147"
misc_feature      7360..8831
                  /note="assembly_name:Contig159"
misc_feature      8932..10179
                  /note="assembly_name:Contig163"
misc_feature      10280..11586
                  /note="assembly_name:Contig165"
misc_feature      11687..12940
                  /note="assembly_name:Contig170"
misc_feature      13041..14154
                  /note="assembly_name:Contig174"
misc_feature      14255..15728
                  /note="assembly_name:Contig177"
misc_feature      15829..17212
                  /note="assembly_name:Contig179"
misc_feature      17313..18620
                  /note="assembly_name:Contig180"

```

Query Match

```

Best Local Similarity 85.5%; Score 18.8; DB 2; Length 317515;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 AAGACTTCTGACTATACATCA 22

DB 17367 AAGACTTCTGACTATACATCA 17388

RESULT 13
AE011043/c

LOCUS	AE011043	10243 bp	DNA	linear	BCT 03-APR-2002
DEFINITION	Methanosarcina acetivorans str. C2A, section 388 of 534 of the complete genome.				
ACCESSION	AE011043 AE010299				
VERSION	AE011043.1 GI:19917413				
KEYWORDS					
SOURCE	Methanosarcina acetivorans C2A.				
ORGANISM	Methanosarcina acetivorans C2A				
REFERENCE	Archaea: Euryarchaeota: Methanococci: Methanosarcinales: Methanosarcinaceae: Methanosarcina.				
AUTHORS	1 (bases 1 to 10243) Galagan, J.E., Nussbaum, C., Roy, A., Endrizzi, M.G., Macdonald, P., Fitzhugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeAvello, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Titrrell, A., Ye, W., Zimmer, A., Barber, R.D., Cann, I., Graham, D.E., Grahame, D.A., Guss, A., Hedderich, R., Ingram-Smith, C., Kuetner, C.H., Krzycki, J.A., Leigh, J.A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J.N., Smith, K., Springer, T.A., Umayam, L.A., White, O., White, R.H., de Macario, E.C., Ferry, J.G., Jarrell, K.F., Jing, H., Macario, A.J.L., Paulsen, I., Pritchett, M., Sowers, K.R., Swanson, R.V., Zinder, S.H., Lander, E., Metcalf, W.W. and Birren, B.				
TITLE	The Genome of <i>M. acetivorans</i> Reveals Extensive Metabolic and Physiological Diversity				
JOURNAL	Genome Res. 12 (4), 532-542 (2002)				
MEDLINE	21929760				
PUBMED	11932238				
REFERENCE	2 (bases 1 to 10243)				
AUTHORS	Birren, B.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-MAR-2002) Center for Genome Research, Whitehead Institute, Nine Cambridge Center, Cambridge, MA 02141, USA				
FEATURES	Location/Qualifiers				
source	1..10243				
gene	/organism="Methanosarcina acetivorans C2A"				
	/strain="C2A"				
	/db_xref="taxon:188937"				
	complement(88..603)				
CDS	/gene="MA3371"				
	complement(88..603)				
	/gene="MA3371"				
	/codon_start=1				
	/transl_table=11				
	/product="predicted protein"				
	/protein_id="AA06740.1"				
	/db_xref="GI:19917414"				
	/translation="MIRESNIMAIMGREGRGRLDIDIGIVVSGIGSNORNOFNGLFSGIDGLYIGPSNPFORTISIGIGVIGSNGSLAKLGIHPRRAGLQMLPPIVYGVSGSFLALPVSSEGAATIPILMVPAAVYQENIKELRQGTETILMLRSYISGRPF"				
gene	complement(859..1488)				
	/gene="MA3372"				
CDS	complement(859..1488)				
	/gene="MA3372"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AA06741.1"				
	/db_xref="GI:19917415"				
	/translation="MKIAVENENQOQSIFEPGTAIVVEEDGVEKKVILNRENOCNARGMAAVMAVGAIDKIDGVKVVVAISEIGISGTPQAGPIFIIVDSKVLVDLILKMDLETERKROEPRKFDTEPLEFRENQDGSINIEDIMEFEPDLSKTLIPYLKNGEENRDIYICGHHKMFVTDLGAMGFEYETVNESTNRKTVRVAQT"				
gene	1981..3642				
	/gene="11VD"				
	/note="MA3373"				
CDS	1981..3642				
	/gene="11VD"				
	/codon_start=1				
	/transl_table=11				
	/product="dihydroxy-acid dehydratase"				
	/protein_id="AA06742.1"				
	/db_xref="GI:19917416"				

gene	/translation="MRSDIIEGPERAPNRSLLKATGVTSEMKKPTIAVVSNDIITPGSMHMLNKLAEAVKAGIRNAGVPEFHHVGVGCDGIAMGHEMKSLSPRTIEPTILAMKAOHODGMVLIPSCDKIVPGCHMAARLDIPAIIVTTCGMLEGIYDNDYRDLISVSEGVAVSAGKISIEPEPKRLNENSCAGASGSCMTANTMCAETALSLPGCATAHAVDAKILILAKESGERIVELVKNLTPKIVTQKSFENAINMDAVGSGTITLPLPAIAHREGDLPLEAFDELRSKTPHLISLRPGGNFLIHAEGIAVLKSLAPGSAKLHLDQILVNGKTIGENLDELEIVNPLNMEITTLLENPIHAEGIAVLKSLAPGSAVVKOAADPKNHVYTPAKVYDCEDEAMKSLIADGVKPGDIIIVIRYRGPGKSGEMLAATAALGRLLSVALVTDGREGSGTRPCIGHVSPEASDGPALIVYDGLIETINPERALNKIIEEELKOKASFVPKPEITGITIARVORAVHSANTGIVD"				
gene	complement(3805..4149)				
	/gene="MA3374"				
CDS	complement(3805..4149)				
	/gene="MA3374"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AA06743.1"				
	/db_xref="GI:19917417"				
	/translation="MNIADVGRKKFNCISFGSROTKQLAMEKKEIGEVMTDHWVAIEFETIHTOSKAEYVVEGNGILRHFLARLRKTKCYTKSLEMKYSVLLMKHKNELSTFN"				
gene	complement(4053..4487)				
	/gene="MA3375"				
CDS	complement(4053..4487)				
	/gene="MA3375"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AA06744.1"				
	/db_xref="GI:19917418"				
	/translation="MNCRCSSNHTKNGIVGQRYKCHDCGYNSVELKSTAPSPLVKQALQYLELGLRIGRLGVSHSVQMKIKFQCEIEELKENSEISIVELDEMHLYIGKKNINAGSGLLIBELGRNSTALLVAERKLDVSGKN"				
gene	4675..4950				
	/gene="MA3376"				
CDS	4675..4950				
	/gene="MA3376"				
	/codon_start=1				
	/transl_table=11				
	/product="predicted protein"				
	/protein_id="AA06745.1"				
	/db_xref="GI:19917419"				
	/translation="MGVIATRLFGGLIKRSQSDVETPTGNGIGYPAATKDIINPATKSVESVDLILGIPETIYNTDVEYTERIERLESQRTREYRPEY"				
gene	5693..7393				
	/gene="MA3377"				
CDS	5693..7393				
	/gene="MA3377"				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AA06746.1"				
	/db_xref="GI:19917420"				
	/translation="MCTTILITRGASMDGSMITVAHSDDELGNRIITYVPAODHROGAMROYFEHYRPRPLVYERGNPERPAHHNSQSPNVPDLTGPIGTIPVRRHYAEDGNGVIMNENHLMGECTNGAKIOPAHVASAEAKNSKHRLTFYSOELSRVALERCKAREAVLEGLKLDKTYISTEETLLVADEDAWFEKCALPDEQYHSAMVAQVDPDEYFVAANERIRIQIEDNRDIIICSELKGLKGLNLMNPEDEPLDMLSSVNGEYNNHYSLRWVVLVLRVNDLALSPMVRGFEVDYDGSYALIDYDEPTKPCPLILOVFLALYDHVGTQFDLIRGVAAAGVDPNRFPGVYQGNQNNYDNEHMYGAMERPISTFYQGYTYVQDLRKTPEIPEASGICLWGPDPVSYTCCTPPAAQALEPERIOSDPVDFPKTYAWAFDFPVSMMARLNTYORRVILIPQOKIETEEVQKTILOWDECCRSKSLERADLITRLGRKANAEVVKQMKRLGHALIARTSDGICNLPNQBSPONIGYSSKWLALTNKDGPTSTEMKP"				
gene	complement(7624..7914)				
	/gene="MA3378"				
CDS	complement(7624..7914)				
	/gene="MA3378"				
	/codon_start=1				
	/transl_table=11				
	/product="predicted protein"				
	/protein_id="AA06747.1"				

gene
CDS
/db_xref="GI:19917421"
/translation="MISMDSMEKIFGKTAQMTVLKMLIENONEPTVLSGIAETGLSH
SSSVRTPLIESGVLEKLGKQIRTFQPMSEFAKLIVFYKINQIVB"
complement(8740..10059)
/gene="MA3379"
/gene="MA3379"
complement(8740..10059)
/codon_start=1
/transl_table=1
/product="amino acid transporter"
/protein_id="AA06748.1"
/db_xref="GI:19917422"
/translation="MEEVKINSNGVNGRESELKRDLSLEITIVIGDILAGIYVM
GKAAGLAGNMWAFLEPFTASLSALSYELSSMTYPRGAETVPARKAFGERIGLTV
GLVIVYVAITASAVAGFGRYSNIFGGILGALSILVLSLVWYGIKSARLAI
LMFTELSELMITIVYGIPIYGVNPEPPLSGVFEASALITFAFGEDIVRLSQE
TKDAKKTTPKRLILSIFETLEIVYICVAVASVLDFOVLGISEIPLAEVAANFGKA
FVLSWIALBETMTVIVVMIGSRIVYGMANGSLPKILARVHOKLTPMTAICGIA
FFSLFVFLGDIATVANIANPMIFIVFYINISLILKRYTPERKRPFPVPSIGRFP
IFPALGALSAVLEFSQIGKEVMLIGFPLAGISALVILKTRKENEL"

BASE COUNT 3037 a 2258 c 2111 g 2837 t

ORIGIN

Query Match 83.6%; Score 18.4; DB 1; Length 10243;
Best Local Similarity 95.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAGACTTCTGACTACATC 20
1 |||||
Db 2740 AGGACTCTGAGTACATC 2721

RESULT 14
AE010860/c 10467 bp DNA linear BCT 03-APR-2002
LOCUS Methanosarcina acetivorans str. C2A, section 205 of 534 of the
DEFINITION complete genome.
ACCESSION AE010860 AE010299
VERSION AE010860.1 GI:1915697
KEYWORDS
SOURCE Methanosarcina acetivorans C2A.
ORGANISM Methanosarcina acetivorans C2A.
REFERENCE 1 (bases 1 to 10467)
AUTHORS Galagan, J.E., Nusbamm, C., Roy, A., Endrizzi, M.G., MacDonald, P.,
Ritzhugh, W., Calvo, S., Engels, R., Smirnov, S., Altmor, D., Brown, A.,
Allen, N., Naylor, J., Stange-Thomann, N., Dearrellano, K., Johnson, R.,
Linton, L., McEwan, P., McKernan, K., Talamas, J., Tittell, A., Ye, W.,
Zimmer, A., Barber, R.D., Cann, I., Graham, D.E., Grahame, D.A.,
Guss, A., Hedderich, R., Ingram-Smith, C., Kuetner, C.H.,
Krzyski, J.A., Leigh, J.A., Li, W., Liu, J., Mukhopadhyay, B.,
Reeve, J.N., Smith, K., Springer, T.A., Umayam, L.A., White, O.,
White, R.H., de Macario, E.C., Perry, J.G., Jarrell, K.F., Jing, H.,
Macario, A.J.L., Paulsen, I., Pritchett, W., Sowers, K.R.,
Swanson, R.V., Zinder, S.H., Lander, E., Melcali, W.W. and Birren, B.
The genome of *M. acetivorans* Reveals Extensive Metabolic and
Physiological Diversity
Genome Res. 12 (4), 532-542 (2002)

JOURNAL 21929760
MEDLINE 11932238
PUBMED 11932238
REFERENCE 2 (bases 1 to 10467)
AUTHORS Birren, B.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Center for Genome Research, Whitehead
Institute, Nine Cambridge Center, Cambridge, MA 02141, USA
FEATURES
source
1. 10467
/organism="Methanosarcina acetivorans C2A"
/strain="C2A"
/db_xref="taxon:188937"
complement(341..1141)
/gene="MA1801"

gene
CDS
complement(341..1141)
/gene="MA1801"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA05207.1"
/db_xref="GI:19915698"
/translation="MYCWNPETLYAFSSAOKSMGLLTKPPLKGNERYLDVCGDGK
LSAFAIKRIPEGSVIGDLSAAMVCPAKNHYPRKDEPPNLSFMDAGNVPFEESEVDI
FSAALWIKTEPKAIEIVLKGLKSLRPGKSLAOGGKNAEVLVLSNMLDEKX
SPYGVNVPFYGTGPEPKMLKNGFSYRAELYSKMTLKGEGLSGMPASTWHP
FTQRPENLKEEFDLAVNFIKLPDMDGVHIMVLELEADIER"
2004..3665
/gene="11VD"
/note="MA1802"
2004..3665
/gene="11VD"
/codon_start=1
/transl_table=1
/product="dihydroxy-acid dehydratase"
/protein_id="AA05208.1"
/db_xref="GI:19915699"
/translation="MRSALITKEGPERANRSLKATGVTDESEKKRPFIAVNSMNEIV
FGHILHDLKADAVKAGIRNAGVPEPFRHTIGVCDGIAMGHEGKYSLEPREVIEDIE
LMVRAHDFDGNVLTPTCKITVPGHLAGRLDIPATVYTGSPMLPGYVNDKXTDLISV
FEGVAGYAGKLSAEALRLNLSACAGSCAGMTANTMACHTEALGSLPGCATAH
AVDAKKVIAKESGERIVALVKNLTTPRKIVTKSENMIMDMAVAGSTNTTLPLA
LAHFGLELPLKTFDELSTRTPLILSRPGGPNMHPFAGVGAENVVORLSKLHL
OLYNGKIGENDELEIVNPKLNAEITITLNDPIHAEGIALVAKSLAPDSVAKQA
AVDRMRHNGPAPVYDCEDAMKSLIAGVSPKAGDIIVIRIEGPKGPGMRMLATA
AIGGGLIEVALITTDGRFSGTRGPGICGIVSPKASGPGIGLVKRGDLEINIPERI
LNKVTLEELERKRAAFVPRKVTGLYAKIQSVISANTGIVD"
4511..4879
/gene="MA1803"
4511..4879
/gene="MA1803"
/codon_start=1
/transl_table=1
/product="predicted protein"
/protein_id="AA05209.1"
/db_xref="GI:19915700"
/translation="MNESEFRELIMVTEFFLIGIFCKRFLVDSYFCKCKRIIPISH
SMDOIOMIDPTAFRRKSRDEPNSSCIYFAITTAAMRNIGSELSYFGVPVPSH
KTNMGVCFILYSHYTPLS"
complement(5796..6758)
/gene="mtxH"
/note="MA1804"
complement(5796..6758)
/gene="mtxH"
/codon_start=1
/transl_table=1
/product="methyltransferase"
/protein_id="AA05210.1"
/db_xref="GI:19915701"
/translation="MRSVDFGVKMKFRQKDELTANAGIKVGGQPEELPTLLAGTIF
YKHEIPEDAAGLDPRIAAKLVNLDVSGDMGNPHMWHIPEITQESTRTIDPLA
EYKSEFLIDSEGVVRAHASYVEIADRAVYNSINISITSEIELAQSQSDIS
IIGGNAMSSIRGEMLEAGLIGELISIDRCGIYKILDPGSIPTMGAGAGIA
LMTITTAKKWGYPTGSGIHNAPSMNMLNOKKRNPLVYKICVCGSTCLQQAAGDF
ILYGPTEAYVFPMAASDIKIAEAVADLIDEPASRNPINLW"
complement(6986..7579)
/gene="mtxA"
/note="MA1805"
complement(6986..7579)
/gene="mtxA"
/codon_start=1
/transl_table=1
/product="methyltransferase"
/protein_id="AA05211.1"
/db_xref="GI:19915702"
/translation="MKTIAENMPVGRGVYTGKPPDSRIAVVTLASHLEAFNNAIMGS
SKTENGVKRTIVNTSNNITVYVCGTESGKAGSLIAIHANGIDKRGIVGSGO
GAIPFIENTISRAVDPRFQROVTLIDRLIGLNDPEETRKIVEDYRQKGEVYPEPVPVCA

Tue Mar 18 16:16:18 2003

us-09-836-439-6.rge

Page 18

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 GACTTCTGAGTACATCAA 22
|||||
Db 29842 GACTTCTGAGTACATCAA 29861

Search completed: March 17, 2003, 11:38:09
Job time : 574.495 secs

CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I) by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC to AKK8/694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AKK54942 to AKK54950 and AKK82169
CC represent sequences used in the exemplification of the present invention.
XX

SQL Sequence 48908 BP; 14764 A; 10631 C; 10740 G; 12773 T; 0 other;

Query Match 83.6%; Score 18.4; DB 22; Length 48908;
Best Local Similarity 95.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 GACTCTGAGTACATCA 22
|||||
DB 3233 GACTCTGAGTACATCA 3252

RESULT 2
AAC42853
ID AAC42853 standard; DNA; 801 BP.
XX
AC AAC42853;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37100.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PR 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-01448814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

```

PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149928.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 10-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160860.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

```

```

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 80.9%; Score 17.8; DB 21; Length 801;
Best Local Similarity 90.5%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 AAGACTTCTGAGTACATCA 21
   ||||| ||||| |||||
DB 543 AAGACTTCTGAGTACATCA 563

```

```

RESULT 3
AA003999/c
ID AA003999 standard; DNA; 877 BP.

```

```

AC AA003999;

```

```

DT 03-SEP-1990 (first entry)

```

```

DE Sequence complementary to dystrophin gene.

```

```

KW X-chromosome; ornithine transcarbamylase deficiency;
    muscular dystrophy; dystrophin; ds.

```

```

OS Synthetic.

```

```

PN EP364255-A.

```

```

XX 18-APR-1990.

```

```

XX 11-OCT-1989; 89EP-0310424.

```

```

XX 12-OCT-1988; 88US-0256689.

```

```

PA (BAYU ) BAYLOR UNIV COLLEGE.

```

```

PI Caskey CT, Chamberlain JS, Gibbs RAL, Rainer JE, Nguyen PN;

```

```

DR WPI: 1990-117752/16.

```

```

PT Multiplex genomic DNA amplification for deletion detection -
    useful for detecting X-linked diseases such as ornithine
    transcarbamylase deficiency and X-linked muscular dystrophy.

```

```

XX Claim 9; Page 19; 32pp; English.

```

```

CC Paired oligonucleotide primers are used in detecting deletions
    specifically of the X and Y chromosomes.

```

```

CC Dystrophin gene may be isolated this way.

```

```

SQ Sequence 877 BP; 289 A; 160 C; 131 G; 297 T; 0 other;

```

```

Query Match 80.9%; Score 17.8; DB 11; Length 877;
Best Local Similarity 90.5%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 AAGACTTCTGAGTACATCA 21
   ||||| ||||| |||||
DB 132 AAGCTTCTGAGTACATCA 112

```

```

RESULT 4
AAC50782

```

```

ID AAC50782 standard; DNA; 1353 BP.

```

```

AC AAC50782;

```

```

DT 18-OCT-2000 (first entry)

```

```

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 66116.

```


PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0156559.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 80.9%; Score 17.8; DB 21; Length 1353;
 Best Local Similarity 90.5%; Pred. No. 59;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGACTTCTGAGTAACATCA 21
 ||||| ||||| ||||| |||||
 Db 543 AAGACTTCTGAGTAACATCA 563

RESULT 5

AAK54111/C
 ID AAK54111 standard; cDNA; 513 BP.

XX AAK54111;

XX 16-NOV-2001 (first entry)

DE Murine transport and binding associated protein encoding cDNA SEQ ID 676.

XX Murine; liver; gene library; amino acid synthesis; binding protein;

KW cell metabolism; energy metabolism; fatty acid metabolism; synthesis;

KW phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;

KW replication; transcription; translation; transport protein; ss.

XX Mus musculus.

XX DE20103510-U1.
 PN 07-JUN-2001.
 XX 28-FEB-2001; 2001DE-2003510.
 PD 02-DEC-1999; 99DE-1058160.
 XX (LION-) LION BIOSCIENCE AG.
 XX WPI; 2001-368570/39.
 DR Gene library containing sequences with specific 3'-ends and no polyA
 PT tail, encoding proteins involved in a wide range of cellular processes
 PT
 PS
 XX
 CC This invention describes a novel gene library (A) comprises a gene
 CC sequence (or its part) encoding a protein involved in amino acid
 CC synthesis, cellular/energy metabolism, metabolism of
 CC fatty acids/phospholipids, synthesis or breakdown of
 CC purines/pyrimidines/nucleosides/nucleotides, DNA
 CC replication/transcription/translation, or is a transport/binding protein.
 CC (A) are produced that correspond to the 3'-end of mRNA but without the
 CC polyA tail. They can be prepared more efficiently and with less effort
 CC than conventional libraries. AAK53436-AAK54275 represent fragments of the
 CC gene library described in the method of the invention.
 XX
 SQ Sequence 513 BP; 143 A; 125 C; 107 G; 137 T; 1 other;

Query Match 78.2%; Score 17.2; DB 22; Length 513;
 Best Local Similarity 86.4%; Pred. No. 1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AAGACTTCTGAGTAACATCA 22
 ||||| ||||| ||||| |||||
 Db 22 AAGACTTCTGAGTAACATCAA 1

RESULT 6

ABL07736
 ID ABL07736 standard; cDNA; 7626 BP.

XX ABL07736;

DE 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 17690.

DE Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB63633.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions.

XX Claim 1; SEQ ID NO 17690; 21bp + Sequence Listing; English.

PS

CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB161737-AB161702).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp://ipr.int/pub/published_pct_sequences.

XX

SO Sequence 7626 BP; 1991 A; 1673 C; 1791 G; 2171 T; 0 other;

Query Match 78.2%; Score 17.2; DB 23; Length 7626;
 Best Local Similarity 86.4%; Pred. No. 1.3e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGACTTCTGAGTACATCA 22
 ||| ||||| ||||| ||
 Db 5862 AAGACTTCTGAGTACATCA 5883

RESULT 7
 AAH68529 standard; DNA; 349980 BP.

XX
 XX AAH68529;
 XX

DT 26-SEP-2001 (first entry)

XX

DE C glutamicum coding sequence fragment SEQ ID NO: 7064.

XX
 XX
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-0127668.
 PT
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI: 2001-376931/40.

XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT

XX
 XX Disclosure: SEQ ID NO: 7064; 246bp + Sequence Listing; English.

XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of *Corynebacterium glutamicum*, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from *Corynebacterium glutamicum*, and identifying a homologue of a gene derived

CC from *Coryneform bacterium*. *Coryneform bacteria* are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX

SO Sequence 349980 BP; 82466 A; 95954 C; 90516 G; 81044 T; 0 other;

Query Match 78.2%; Score 17.2; DB 22; Length 349980;
 Best Local Similarity 86.4%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGACTTCTGAGTACATCA 22
 ||| ||||| ||||| ||
 Db 232860 AAGACTTCTGAGTACATCA 232881

RESULT 8
 AAC57234 standard; DNA; 461 BP.

XX
 XX AAC57234;
 XX

DT 25-JAN-2001 (first entry)

XX

DE *Pinus radiata* transcription factor DNA sequence #631.

XX
 XX
 XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox;
 XX homeodomain zipper; LIM domain; AP2; ERBBs; zinc finger domain;
 XX type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 OS *Pinus radiata*.
 XX
 PN WO200053724-A2.
 XX
 PD 14-SEP-2000.
 XX
 XX 09-MAR-2000; 2000WO-US06112.
 PF
 XX 11-MAR-1999; 99US-0266513.
 PR 18-AUG-1999; 99US-0149485.
 PR
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 PI
 XX WPI: 2000-579369/54.

XX
 XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT

XX
 XX Claim 1; Page 624; 747bp; English.

XX
 XX The present invention relates to novel plant transcription factors from
 CC *Eucalyptus grandis* or *Pinus radiata*. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

Sequence 461 BP; 163 A; 97 C; 94 G; 107 T; 0 other;
Query Match 76.4%; Score 16.8; DB 21; Length 461;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 GACTCTGAGTACATCAA 22
||||| ||||| |||
Db 408 GACTTCTCAGTACAGTCAA 427
RESULT 9
AAK66626
ID AAK66626 standard; DNA; 21477 BP.
XX AAK66626;
AC AAK66626;
XX 06-NOV-2001 (first entry)
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21438.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 15-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-020515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.

PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 26-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249247.
 PR 17-NOV-2000; 2000US-0249251.
 PR 17-NOV-2000; 2000US-0249252.
 PR 17-NOV-2000; 2000US-0249253.
 PR 17-NOV-2000; 2000US-0249254.
 PR 17-NOV-2000; 2000US-0249255.
 PR 17-NOV-2000; 2000US-0249256.
 PR 17-NOV-2000; 2000US-0249257.
 PR 17-NOV-2000; 2000US-0249258.
 PR 17-NOV-2000; 2000US-0249259.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0250393.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPL; 2001-483426/52.
 XX
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 PS
 XX Disclosure; SEQ ID NO 21437; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 CC
 SQ Sequence 21480 BP; 5315 A; 5005 C; 5257 G; 5903 T; 0 other;
 XX
 Query Match 76.4%; Score 16.8; DB 22; Length 21480;
 Best Local Similarity 90.0%; Pred. No. 2.3e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 DB 7222 AGCCTTCGTGATGACCAATCA 7241
 QY 2 AGACCTTCGTGATGACCAATCA 21
 II |||||
 Db 7222 AGCCTTCGTGATGACCAATCA 7241
 RESULT 11
 AAS01375/C
 ID AAS01375 standard; CDNA; 3114 BP.
 AC AAS01375;
 XX
 DT 04-JUL-2001 (first entry)
 XX
 DE Human TANGO 405 CDNA sequence.
 XX
 KW Human; TANGO 210; clone jthla152h06; TANGO 364; TANGO 366; dectin-2;
 KW INTERCEPT 394; INTERCEPT 400; TANGO 405; cellular process regulator;
 KW gene therapy; growth modulator; proliferation; cell differentiation;
 KW lymphocyte; bone marrow cell migration; leukaemia; lymphoma;
 KW autoimmune disorder; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 154..783
 FT /tag= a
 FT /product= "TANGO 405 protein"
 FT /note= "The ORF is specifically claimed"
 FT sig_peptide 154..297
 FT /tag= b
 FT mat_peptide 298..780
 FT /tag= c
 XX
 PN WC200118016-A1.

XX 15-MAR-2001.
PD 30-JUN-2000; 2000MO-US18174.
XX 10-SEP-1999; 99US-0393996.
PR (MILL-) MILLENNIUM PHARM INC.
XX
PI Fraser CC, Sharp JD, Wighton N, Myers PS, Goodearl ADJ;
XX WPI: 2001-183280/18.
DR P-PSDB: AAU00479.
XX
PT Isolated nucleic acid molecules encoding proteins useful as modulating
PT agents in regulating a variety of cellular processes are used for
XX treating e.g. cancer and autoimmune disorders.
PS Claim 2: Fig 6A-6C; 326pp; English.
XX
CC The present sequence encoding for human TANGO 405 protein is isolated
CC from cDNA clone jh1a15206 from a human mixed lymphocyte reaction cDNA
CC library. It is 1 of 6 novel human proteins which include TANGO 210
CC (AAU00469), TANGO 364 (AAU00471), TANGO 366 (AAU00472), INTERCEPT 394
CC (AAU00473), and INTERCEPT 400 (AAU00476). Novel sequences for murine
CC TANGO 210 (AAU00470), INTERCEPT 400 (AAU00477), TANGO 405 (AAU00480) and
CC a rat INTERCEPT 400 (AAU00478) sequence are also described. The nucleic
CC acids encoding these novel proteins are useful as modulating agents in
CC regulating a variety of cellular processes and can be used to express
CC the proteins in a host cell in gene therapy applications. Human and
CC murine TANGO 405 proteins show sequence homology to murine decalin-2,
CC TANGO 405 modulates growth, proliferation, survival, differentiation,
CC activity, morphology and movement/migration of human lymphocytes and
CC bone marrow cells and tissues and can be used to prevent, diagnose or
CC treat leukemia, lymphomas and autoimmune disorders.
SQ
Sequence 3114 BP; 1001 A; 527 C; 517 G; 1069 T; 0 other;
Query Match 74.5%; Score 16.4; DB 22; Length 3114;
Best Local Similarity 94.4%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 5 CTTCTGAGTAACATCA 22
|||||||
DB 1885 CTTCTGAGTAACATCA 1868
RESURF 12
ID ABA18026
AC ABA18026 standard; DNA; 29329 BP.
XX
AC ABA18026;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 10357.
XX
KW Human; nootropic; neuroprotective; cytoskeletal; dermatological; virologic;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisticking; antianaemic; antithrombotic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01334.
XX

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228928.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 25-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236570.
PR 02-OCT-2000; 2000US-0236802.

xx	Disclosure; SEQ ID NO 10357; 1701tp + Sequence Listing; English.
xx	
cc	The invention relates to novel genes (ABA11004-ABA21534) and proteins
cc	(ABA14678-ABA18001) useful for preventing, treating or ameliorating
cc	medical conditions e.g. by protein or gene therapy. The genes are
cc	isolated from a range of human tissues disclosed in the specification.
cc	The nucleic acids, proteins, antibodies and (ant)agonists are useful
cc	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
cc	and ovarian cancer and other cancers of the adrenal gland, bone, bone
cc	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
cc	(b) immune disorders e.g. Addison's disease, diabetes mellitus, Crohn's
cc	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, autoimmune
cc	disease, multiple sclerosis, rheumatoid arthritis and ulcerative
cc	colitis; (c) cardiovascular disorders such as myocardial ischaemias;
cc	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
cc	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
cc	and parasitic infections.
cc	Note: The sequence data for this patent did not form part of the
cc	printed specification, but was obtained in electronic format directly
cc	from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
xx	
sq	Sequence 29329 BP; 6864 A; 6756 G; 7225 G; 8484 T; 0 other:
	Query Match 74.5%; Score 16.4; DB 22; Length 29329;
	Best Local Similarity 94.4%; Pred. No. 3.6e+02;
	Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0
QY	3 GACTTCGAGTACAAATC 20
Db	8110 GACTTCGAGAACAAATC 8127
RESULT 13	
ID	ABA20511 standard; DNA; 29329 BP.
xx	
AC	ABA20511;
xx	
DT	23-JAN-2002 (first entry)
xx	
DE	Human nervous system related polynucleotide SEQ ID NO 12842.
xx	
xx	Human; nocotropic; neuroprotective; cytosolic; dermatological; virucide;
xx	immunosuppressive; antilinfimatoxy; anti-HIV; antibacterial; vulnereary;
xx	antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
xx	antipneumatic; hepatotropic; cerebroprotective; antiinflammatory;
xx	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
xx	antiparasitic; cardiac; immune disorder; cardiovascular disorder;
xx	neurological disease; infection; neurotropic; gene therapy; vaccine; ds.
xx	
OS	Homo sapiens.
xx	
PN	WO200159063-A2.
xx	
PD	16-AUG-2001.
xx	
PF	17-JAN-2001; 2001WO-US01334.
xx	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0186874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.

CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 29329 BP: 6864 A; 6756 C; 7225 G; 8484 T; 0 other:

Query Match 74.5%; Score 16.4; DB 22; Length 29329;
Best Local Similarity 94.4%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 3 GACTCTGTGATACATC 20
DB 8110 GACTCTGTGATACATC 8127

RESULT 14
AAK70791
ID AAK70791 standard; DNA: 29329 BP.
XX
AC AAK70791;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25603.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225470.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230458.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.

PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 06-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2000US-0259678.
 PR XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 PT WPI: 2001-483426/52.
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 PS
 XX Disclosure: SEQ ID NO 33324; 3071pp + Sequence Listing; English.
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent
 CC diagnosis and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 CC
 CC
 CC
 XX
 SO Sequence 29329 BP; 6864 A; 6756 C; 7225 G; 8484 T; 0 other;

QY 3 GACTTCTGAGTACAATC 20
 ||||||||| |||||||
 Db 8110 GACTTCTGAGTACAATC 8127

Search completed: March 17, 2003, 10:52:21
 Job time : 217.253 secs

Query Match 74.5%; Score 16.4; DB 22; Length 29329;
 Best Local Similarity 94.4%; Pred. NO.3.6e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2003, 10:23:12 ; Search time 777.688 Seconds
(without alignments)
458.154 Million cell updates/sec

Title: US-09-836-439-6

Sequence: 1 aagactctgagtaacatcaa 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estfun:*
17: gb_est7:*
18: em_gss_hum:*
19: em_gss_hum:*
20: em_gss_hum:*
21: em_gss_vrt:*
22: em_gss_vrt:*
23: em_gss_vrt:*
24: em_gss_vrt:*
25: em_gss_vrt:*
26: em_gss_vrt:*
27: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.4	83.6	347	9	AA970603 op40f09.s
C 2	18.4	83.6	395	9	AA837289 op40f09.s
C 3	17.8	80.9	364	9	AA229942 nc51h06.r
C 4	17.8	80.9	418	14	B0622226 rchic.pk0
C 5	17.8	80.9	509	12	B0882650 sae94g07.
C 6	17.8	80.9	545	9	AT738689 w122a08.x

Result No.	Score	Query Match	Length	DB ID	Description
C 7	17.8	80.9	580	9	AA546700 vk55a10.r
C 8	17.8	80.9	857	17	BH138512 ENTN176TF
C 9	17.4	79.1	375	12	BF609959 NXSI_053
C 10	17.4	79.1	474	10	AM699516 gpb06f02.y
C 11	17.4	79.1	477	12	BG040453 NXSI_108
C 12	17.4	79.1	549	10	AW738913 gbl6b06.y
C 13	17.4	79.1	608	17	AQ525332 HS_5226.B
C 14	17.4	79.1	611	13	BJ173371 BJ173371
C 15	17.4	79.1	692	13	BJ171523 BJ171523
C 16	17.4	79.1	734	13	BJ167340 BJ167340
C 17	17.2	78.2	106	10	BE339863 EST343923
C 18	17.2	78.2	185	17	AZ242233 RPT-23-7
C 19	17.2	78.2	236	10	BB528073 BB528073
C 20	17.2	78.2	258	10	BB301743 BB301743
C 21	17.2	78.2	325	17	AZ746982 RPT-24-1
C 22	17.2	78.2	328	13	B1932924 EST552813
C 23	17.2	78.2	390	12	BF451198 u266c12.x
C 24	17.2	78.2	391	12	BF333029 msa38b04.
C 25	17.2	78.2	396	17	AZ104881 RPT-23-4
C 26	17.2	78.2	409	12	BE851242 uv93h01.y
C 27	17.2	78.2	417	12	BE848177 uv36b05.y
C 28	17.2	78.2	420	14	BM899883 ut-m-dj1-
C 29	17.2	78.2	425	9	A1663816 u106c05.x
C 30	17.2	78.2	432	13	BI474888 t554811.x
C 31	17.2	78.2	436	9	AU014847 AU014847
C 32	17.2	78.2	450	14	BQ132786 t68g09.x
C 33	17.2	78.2	454	17	AQ769890 HS_3084.B
C 34	17.2	78.2	460	10	BE579476 kg28d05.y
C 35	17.2	78.2	473	10	AM220604 EST296989
C 36	17.2	78.2	482	9	A1267625 ag92b06.x
C 37	17.2	78.2	482	10	AM398852 EST309352
C 38	17.2	78.2	487	10	BE462943 EST324995
C 39	17.2	78.2	489	9	AU080451 AU080451
C 40	17.2	78.2	489	9	AA270973 va84c09.r
C 41	17.2	78.2	492	12	BF116785 u202c01.y
C 42	17.2	78.2	511	12	BE016374 uy40b09.y
C 43	17.2	78.2	512	10	BE689588 uw56d01.y
C 44	17.2	78.2	515	12	BG359610 sac26d02.
C 45	17.2	78.2	518	17	AQ416223 RPT-11-1

ALIGNMENTS

RESULT 1
AA970603/c 347 bp mRNA linear EST 07-JUL-1998
LOCUS op40f09.s1 Soares_NFL.T.GRC.S1 Homo sapiens CDNA clone
DEFINITION IMAGE:1579337 3', mRNA sequence.
ACCESSION AA970603
VERSION AA970603.1 GI:3145110
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 347)
NCT-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LML ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1053 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 344.
Location/Qualifiers
1. 347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1579337"

```

/c/clone_11b="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled: Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker: Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDHL19W, Testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT      91 a      75 c      63 g      118 t
ORIGIN

Query Match      83.6%; Score 18.4; DB 9; Length 347;
Best Local Similarity 95.0%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      3 GACTTCTGAGTACCAATCAA 22
        |||||||||||
Db      335 GACTTCTGAGTACCAATCAA 316

RESULT 2
AA837289/c      395 bp      mRNA      linear      EST 31-MAR-1998
LOCUS      od66e07.s1 NCI-CGAP_GCB1 Homo sapiens CDNA clone IMAGE:1369092,
DEFINITION      mRNA sequence.
ACCESSION      AA837289
VERSION      AA837289.1 GI:2912488
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 395)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 1043 Std Error: 0.00
Seq primer: -40m13 fwd. EF from Amersham
High quality sequence stop: 336.
Location/Qualifiers
1. 395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1369092"
/clone_11b="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAGTGGAGCGGCGCTCATTTTCTTTTCTTTT-3'

```

```

]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      100 a      77 c      73 g      145 t
ORIGIN

Query Match      83.6%; Score 18.4; DB 9; Length 395;
Best Local Similarity 95.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      3 GACTTCTGAGTACCAATCAA 22
        |||||||||||
Db      345 GACTTCTGAGTACCAATCAA 326

RESULT 3
AA229942/c      364 bp      mRNA      linear      EST 21-AUG-1997
LOCUS      nc51h06.r1 NCI-CGAP_Pr3 Homo sapiens CDNA clone IMAGE:1011707
DEFINITION      similar to contains Alu repetitive element, mRNA sequence.
ACCESSION      AA229942
VERSION      AA229942.1 GI:1852255
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 364)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui, M.D.
Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 616 Std Error: 0.00
Seq primer: -28m13 rev1 EF from Amersham.
Location/Qualifiers
1. 364
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1011707"
/clone_11b="NCI-CGAP_Pr3"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected cells histologically-determined to be
fully malignant prostate cancer cells. Double-stranded
cDNA was ligated to EcoRI adaptors, 5 cycles of PCR
applied to the cDNA with an adaptor-specific primer, and
the resulting PCR product subcloned into PAMP10 by the
UDC-cloning method (Life Technologies). Average insert
size is 600 bp. NOTE: Not directionally cloned. This
library was constructed by David Krizman."
BASE COUNT      101 a      82 c      86 g      95 t
ORIGIN

Query Match      80.9%; Score 17.8; DB 9; Length 364;
Best Local Similarity 90.5%; Pred. No. 7.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 2 AGACTTCTGAGTAACATCAA 22
 Db 48 AGCTCTGCTGAGTAACATCAA 28

RESULT 4

LOCUS B06222226

DEFINITION cDNA, B06222226 418 bp mRNA linear EST 01-JUL-2002

ACCESSION B06222226

VERSION B06222226.1 GI:21649395

KEYWORDS

EST

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 418)
 EST analysis of genes expressed by the zygomycete pathogen

JOURNAL

Unpublished (2002)

COMMENT

Contact: Freiloser F. M.
 Department of Entomology
 University of Maryland
 4112 Plant Sciences Building, College Park, MD 20742, USA
 Tel: 301 405 16 13
 Fax: 301 314 92 90
 Email: f34@umail.umd.edu

FEATURES

SOURCE

1. .418
 Location/Qualifiers

/organism="Conidiobolus coronatus"

/strain="ARSEF 512"

/db_xref="taxon:34488"

/clone_id="Conidiobolus coronatus ARSEF 512"

/note="Vector: Unizap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and the unidirectional Lambda vector Unizap."

BASE COUNT 119 a 85 c 76 g 136 t

2 others

Query Match

Best Local Similarity 90.9%; Score 17.8; DB 14; Length 418;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 AGACTTCTGAGTAACATCAA 22

|||||

Db 175 AGACTTCTGAGTAACATCAA 195

|||||

RESULT 5

LOCUS

BG882650

DEFINITION

sa94907.y1 Gm-cl065 GLYCINE max cDNA clone GENOME SYSTEMS CLONE

ACCESSION

BG882650

VERSION

BG882650.1 GI:14259742

KEYWORDS

EST

SOURCE

soybean

ORGANISM

Glycine max

REFERENCE

1 (bases 1 to 509)
 Shoemaker R., Kelm P., Vodkin L., Erpelting J., Coryell V., Khanna A., Bolla B., Merritt M., Hillier L., Kucaba T., Martin J., Beck C.,

AUTHORS

Wyllie T., Underwood K., Steptoe M., Thelshing B., Allen M., Bowers Y., Pearson B., Swaller T., Gibbons M., Pape D., Harvey N., Schurk R., Ritter E., Kohn S., Shih T., Jackson Y., Cardenas M., McCann R., Waterston R. and Wilson R.

TITLE

Public Soybean EST Project

JOURNAL

Unpublished (1999)
 Contact: Shoemaker R./Public Soybean EST Project

COMMENT

Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu

FEATURES

This clone is available through: Reggen, Invitrogen Corp. 2130

SOURCE

South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: cut@reggen.com

High quality sequence stop: 415.

Location/Qualifiers

1. .509

ORGANISM

/organism="Glycine max"

db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl065-3758"

clone_id="Gm-cl065"

/issue_type="germinating shoots"

lab_host="Dhl08"

/note="Vector: Bluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from germinating shoots of the cultivar Williams. The seeds were allowed to germinate for 24 hours prior to being cold stressed for 2 days at 4C. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the plasmid vector. The ligated cDNA fragments were transformed into Dhl08 host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT

151 a 91 c 121 g 146 t

ORIGIN

Query Match

Best Local Similarity 90.9%; Score 17.8; DB 12; Length 509;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 AGACTTCTGAGTAACATCAA 22

|||||

Db 77 AGCTTCTGAGTAACATCAA 57

|||||

RESULT 6

LOCUS

A1738689

DEFINITION

w122a08.x1 NCI-CGAP Col6 Homo sapiens cDNA clone IMAGE:2390966 3'

ACCESSION

A1738689

VERSION

A1738689.1 GI:5100670

KEYWORDS

EST

SOURCE

human

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 545)
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS

NCI-CGAP http://www.nci.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

JOURNAL

Contact: Robert Strussberg, Ph.D.
 Email: cga@bbs.treball.nih.gov

COMMENT

Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.lnl.gov/db/brp/image/image.html

Insert Length: 850 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 446.
Location/Qualifiers

FEATURES

Source

1. 545
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:230966"
/clone_id="NCI.CGAP.C016"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: p773D-pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; plasmid DNA from the normalized library NCI.CGAP.C010 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1057416-1061255, and 114584-114555)."
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT

165 a 131 c 82 g 167 t

Query Match 80.9%; Score 17.8; DB 9; Length 545;
Best Local Similarity 90.5%; Pred. No. 9.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGACTTGTGATACATCA 21
|||||
Db 226 AAGACTTGTGATACATCA 246

RESULT 7
AA546700 580 bp mRNA linear EST 05-AUG-1997
LOCUS V55510.1 StrataGene mouse Tcell 937311 Mus musculus cDNA clone
DEFINITION IMAGE:958554.5, mRNA sequence.
ACCESSION AA546700
VERSION AA546700.1 GI:2307991
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Mammalia: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 580)
Marrin, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Materon, R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:547346
High quality sequence stop: 405.
Location/Qualifiers

FEATURES

Source

1. 580
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:958554"
/clone_id="StrataGene mouse Tcell 937311"
/tissue_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: blood; Vector: pBluescript SK-; Site_1:

ECORI: Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT, M30 CD4+ cells. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3' "

BASE COUNT

162 a 117 c 140 g 161 t

Query Match 80.9%; Score 17.8; DB 9; Length 580;
Best Local Similarity 90.5%; Pred. No. 9.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 AAGACTTGTGATACATCAA 22
|||||
Db 506 AAGACTTGTGATACATCAA 486

RESULT 8
BH138512 857 bp DNA linear GSS 07-AUG-2001
LOCUS ENTN176TF Entamoeba histolytica sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION BH138512.1 GI:15097573
VERSION BH138512
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Eukaryota; Entamoebidae; Entamoeba.

REFERENCE
AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: M13-Forward
Class: shotgun
High quality sequence stop: 8
High quality sequence stop: 461.
Location/Qualifiers

1. 857
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_id="Entamoeba histolytica sheared DNA"
/note="Vector: PROS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

FEATURES

Source

BASE COUNT

276 a 151 c 101 g 323 t

Query Match 80.9%; Score 17.8; DB 17; Length 857;
Best Local Similarity 90.5%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGACTTGTGATACATCA 21
|||||

Db 509 AAGACTCTTGTAACAATCA 489

RESULT 9
LOCUS BFG09959 375 bp mRNA linear EST 14-DEC-2000
DEFINITION NXSI_053_A10_F NXSI (Nsf Xylem Side wood Inclined) Pinus taeda cDNA
clone NXSI_053_A10 5', mRNA sequence.
BFG09959
ACCESSION BFG09959
VERSION BFG09959
KEYWORDS EST.
SOURCE 1lobolly pine.
ORGANISM Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 375)
AUTHORS Sederoff, R.
TITLE Molecular Basis of Wood Formation in the Pine Megagenome
JOURNAL Unpublished (2000)
COMMENT Contact: Johnson, Arthur
North Carolina State University
Tel: 919 515 7800
Fax: 919 515 7801
Email: rjohnson@unc.edu
Seq primer: 73.
Location/Qualifiers
1..375
/organism="Pinus taeda"
/strain="Coastal plain loblolly pine from North Carolina"
/db_xref="taxon:3352"
/clone="NXSI_053_A10"
/clone_1lb="NXSI_053_A10"
/tissue_type="xylem"
/cell_type="side"
/dev_stage="juvenile"
/lab_host="XLI-Blue"
/note="Vector: Bluescript SK; Site 1: Eco RI; Site 2: XhoI
: The library is from early (spring) wood, taken from
three six-year old trees (three different genotypes), in
the juvenile phase. These trees were induced to form side
wood by bending to a 45 degree angle and tying them to the
ground. Differentiating xylem was harvested from the sides
of the inclined stems, and a mixture of all three
genotypes was used for the library. oligo-dt primed cDNA
was directionally cloned into the EcoRI-XhoI Bluescript SK
vector arms. NOTE: The sequences contain a 'cDNA adapter'
between the EcoRI site and the start of the EST. The
adapter sequence is 'AATTCGACGAG'."

BASE COUNT 80 a 76 c 76 g 129 t 14 others
ORIGIN

Query Match 79.1%; Score 17.4; DB 12; Length 375;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACTTCGAGTACAAATCA 22
|||||
Db 132 ACTTCTGAGTACAAACCA 150

RESULT 10
LOCUS AM699516/c 474 bp mRNA linear EST 18-APR-2000
DEFINITION gdb6f02.y1 Moss EST library PPN Physcomitrella patens cDNA clone
PEP_SOURCE_ID: PPN080304 5' similar to TR:09538 Q96538 ACYL-COA
SYNTHETASE; mRNA sequence.
AM699516
VERSION AM699516
KEYWORDS EST.
SOURCE Physcomitrella patens.
ORGANISM Physcomitrella patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

REFERENCE 1 (bases 1 to 474)
AUTHORS Quatrano, R., Bashlades, S., Cove, D., Cumming, A., Knight, C., Clifton
Marra, M., Hillier, L., Page, D., Martin, J., Wylie, T., Underwood
R., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,
Stepcoe, M., Gibbons, M., Harvey, N., Rafter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Leeds/Mash U Moss EST Project
Unpublished (1999)
COMMENT Contact: Ralph Quatrano
Leeds/Mash U Moss EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Libraries were constructed by Dr. Stavros Bashlades as part of the
Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
Seq primer: -40RP from Gibco
High quality sequence stop: 405.
Location/Qualifiers
1..474
/organism="Physcomitrella patens"
/db_xref="taxon:3218"
/clone="PEP_SOURCE_ID: PPN080304"
/clone_1lb="Moss EST library PPN"
/tissue_type="protonemata: 7 day old tissue auxin treated"
/lab_host="DH10B"
/note="Vector: Bluescript SK; Site 1: EcoRI; Site 2:
XhoI: Construction of the cDNA library was carried out
using Stratagene's 'Unizap - cDNA synthesis kit'. cDNA was
constructed using an oligo dt primer/linker that contains
a XhoI site within it. Following ds cDNA synthesis,
EcoRI adapters were ligated to the blunt ends and sample
sticky end on one side and a XhoI sticky end on the other.
This cDNA was ligated directionally in Unizap arms. The
vector is designed containing the pluescript sequence as
well as lambda DNA and cDNA is cloned within this
pluescript sequence. The vector was then packaged using
Gold giga packaging extracts. Library was grown in XL1blue
MRP cells and amplified. The library was excised by mass
excision using Stratagene's 'Mass excision kit' that uses
exonuclease as a helper phage that releases the pluescript
sequence and circularises it as single stranded plasmids
that are then packaged (by helper phage) and secreted out
of the host cell as phagemids. SOLR cells were transformed
with phagemids and the library was plated out on LB-amp
plates to select for transformants. Approximately 1,000
,000 colonies were grown and recovered. The double
stranded plasmid library was recovered by using Qulagen
Midl prep kit. 2 micro grams of each library were used to
transform DH10B cells by electroporation."

BASE COUNT 138 a 99 c 114 g 123 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 10; Length 474;
Best Local Similarity 94.7%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACTTCTGAGTACAAATCA 22
|||||
Db 432 ACTTCTGAGTACAAATCA 414

RESULT 11
LOCUS BG040453 477 bp mRNA linear EST 24-JAN-2001
DEFINITION NXSI_108_D07_F NXSI (Nsf Xylem Side wood Inclined) Pinus taeda cDNA
clone NXSI_108_D07 5', mRNA sequence.
BG040453
ACCESSION BG040453

VERSION BG040453.1 GI:12483038
 KEYWORDS EST.
 SOURCE loblolly pine.
 ORGANISM Pinus taeda
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
 REFERENCE 1 (bases 1 to 477)
 AUTHORS Sederoff, R.
 TITLE Molecular Basis of Wood Formation in the Pine Megagenome
 JOURNAL Unpublished (2000)
 COMMENT Contact: Johnson, Arthur
 North Carolina State University
 Tel: 919 515 7800
 Fax: 919 515 7801
 Email: ajohnson@unity.ncsu.edu
 Seq primer: T3.
 FEATURES
 source
 Location/Qualifiers
 1..477
 /organism="Pinus taeda"
 /strain="Coastal plain loblolly pine from North Carolina"
 /db_xref="taxon:3352"
 /clone="NXSL108.D07"
 /clone_lib="NXSL (NSF Xylem Side wood Inclined)"
 /tissue_type="Xylem"
 /cell_type="Side"
 /dev_stage="Juvenile"
 /lab_host="XLI-Blue"
 /note="Vector: Bluescript SK; Site.1: Eco RI; Site.2: XhoI
 ; The library is from early (spring) wood, taken from
 three six-year old trees (three different genotypes), in
 the juvenile phase. These trees were induced to form side
 wood by bending to a 45 degree angle and tying them to the
 ground. Differentiating xylem was harvested from the sides
 of the inclined stems, and a mixture of all three
 genotypes was used for the library. oligo-dT primed cDNA
 was directionally cloned into the EcoRI-XhoI Bluescript SK
 vector arms. NOTE: The sequences contain a 'cDNA adapter'
 between the EcoRI site and the start of the EST. The
 adapter sequence is 'ATTTCGCGACGAG'."
 BASE COUNT 106 a 106 c 93 g 154 t 18 others
 ORIGIN
 Query Match 79.1%; Score 17.4; DB 12; Length 477;
 Best Local Similarity 94.7%; Pred. No. 1.3e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 4 ACTTCGAGTAACATCAA 22
 ||||||||||||||||
 Db 134 ACTTCGAGTAACATCAA 152
 RESULT 12 549 bp mRNA linear EST 25-APR-2000
 AM738913
 LOCUS GB1606.y1 Moss EST library PPN Physcomitrella patens cDNA clone
 DEFINITION PEP_SOURCE_ID:PPN100912 5' similar to TR:Q96538 ACTYL-COA
 SYNTHETASE ; mRNA sequence.
 ACCESSION AM738913
 VERSION AM738913.1 GI:7647930
 KEYWORDS EST.
 SOURCE Physcomitrella patens.
 ORGANISM Physcomitrella patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
 Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 549)
 AUTHORS Quatrano, R., Bashlars, S., Core, D., Cuming, A., Knight, C., Clifton
 S., Maitra, M., Hillier, L., Pape, D., Martin, V., Wylie, T., Underwood
 K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,
 Stepoe, M., Gibbons, M., Harvey, N., Ritzer, E., Jackson, Y., McCann, R.,
 Waterston, R. and Wilson, R.
 Leeds/Mash U Moss EST Project
 Unpublished (1999)
 COMMENT Contact: Ralph Quatrano

Leeds/Mash U Moss EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: este@wustl.edu
 Libraries were constructed by Dr. Stavros Bashlars as part of the
 Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
 Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
 University Genome Sequencing Center For information on obtaining a
 clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
 Seq primer: -40RP from Glbco
 High quality sequence stop: 408.
 FEATURES
 source
 Location/Qualifiers
 1..549
 /organism="Physcomitrella patens"
 /db_xref="taxon:3318"
 /clone="PEP_SOURCE_ID:PPN100912"
 /clone_lib="Moss EST library PPN"
 /tissue_type="Protonemata: 7 day old tissue auxin treated"
 /lab_host="DH10B"
 /note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
 XhoI; Construction of the cDNA library was carried out
 using Stratagene's 'Unizap - cDNA synthesis kit'. cDNA was
 constructed using an oligo-dT primer/linker that contains
 a XhoI site within it. Following ds cDNA synthesis,
 EcoRI adapters were ligated to the blunt ends and sample
 was digested with XhoI. The result is cDNA with an EcoRI
 sticky end on one side and a XhoI sticky end on the other.
 This cDNA was ligated directionally in Unizap arms. The
 vector is designed containing the pBluescript sequence as
 well as lambda DNA and cDNA is cloned within this
 pBluescript sequence. The vector was then packaged using
 Gold gappackaging extracts. Library was grown in XL1Blue
 MRF' cells and amplified. The library was excised by mass
 excision using Stratagene's 'Mass excision kit' that uses
 exasist as a helper phage that releases the pBluescript
 sequence and circularises it as single stranded plasmids
 that are then packaged (by helper phage) and secreted out
 of the host cell as phagemids. SOLR cells were transformed
 with phagemids and the library was plated out on LB-amp
 plates to select for transformants. Approximately 1,000
 ,000 colonies were grown and recovered. The double
 stranded plasmid library was recovered by using Qugen
 Midi prep kit. 2 micro grams of each library were used to
 transform DH10B cells by electroporation."
 BASE COUNT 151 a 116 c 125 g 157 t
 ORIGIN
 Query Match 79.1%; Score 17.4; DB 10; Length 549;
 Best Local Similarity 94.7%; Pred. No. 1.4e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 4 ACTTCGAGTAACATCAA 22
 ||||||||||||||||
 Db 418 ACTTCGAGTAACATCAA 400
 RESULT 13 608 bp DNA linear GSS 11-MAY-1999
 AO525332
 LOCUS HS_5226_B1-C08-T7A RPT-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate-802 Col-15 Row-F, DNA sequence.
 ACCESSION AO525332
 VERSION AO525332.1 GI:4772652
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 608)
 AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and

TITLE	Hood, L.
JOURNAL	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
COMMENT	99380589
	Contact: Mahatras GG, Wallace JC, Hood L

Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
993805689
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RCGI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.bufralo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufralo.edu/ordering_bac.htm or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 802 row: F column: 15
Seq primer: T7
Class: BAC ends
High quality sequence stop: 608.

FEATURES
SOURCE

Location/Qualifiers
1. .608

BASE COUNT	221 a	123 c	107 g	147 t	10 others
ORIGIN					

Query Match	79.18;	Score 17.4;	DB 17;	Length 608;
Best Local Similarity	85.78;	Pred. No. 1.4e+03;		
Matches 18; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY 1 AAGACTTCTGAGTAACAATCA 21
|| |||| | |||||
Db 22 AANACTTTTNACTAACAATCA 42

RESULT 14				
BJ173371				
LOCUS	BJ173371	611 bp	mRNA	linear
DEFINITION	BJ173371 full length cDNA library, chloromemata and young		EST 24-JAN-2002	

gametophores *Physcomitrella patens* subsp. *patens* cDNA clone
pph35g12 3', mRNA sequence.
R1173371

ACCESSION	BU173371	
VERSION	BU173371.1	GI:18341336
REVISIONS	1	

KEYWORDS	EST.
SOURCE	Physcomitrella patens subsp. patens.
ORGANISM	Physcomitrella patens subsp. patens

REFERENCE
1 (bases 1 to 611)
Fujita, T., Shin-I, T., Seki, M., Kamiya, A., Uchiyama, I., Nishiyama T. and Nakamura, T. 1990. *Journal of Virology*, 64: 1041-1045.

TITLE	Comparison of the moss <i>Physcomitrella patens</i> genome with flowering plants genome
JOURNAL	Unpublished (2002)
COMMENT	Contact: Tadasu Shin-i

Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshn1@genes.nig.ac.jp

A backbone of the vector is basically from pBluescript(KS), that was *in vivo* excised from a modified λ phage vector (ϕ 01 Tec, Germany). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector. cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated basically according to the method described in The Plant J 15, 707-720 (1998) Seki M. et al. Proteome data were blended by the POLYTRON[®], and then cultivated on the BODAG medium for 13-14 days under the continuous light.

FEATURES	Location/Qualifiers
SOURCE	1..611

FEATURES
SOURCE

Location/Qualifiers
1. .611

BASE COUNT	ORIGIN
178 a	135 c 128 g 170 t

Query Match	79.18;	Score 17.4;	DB 13;	Length 611;
Best Local Similarity	94.78;	Pred. No. 1.4e+03;		
Matches 18;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0

OY 4 ACTTCTGAGTAACAATCAA 22
 |||||
 Db 181 ACTTCTGAGTAACCATCAA 199

RESULT 15
BJ171523

LOCUS	692 bp	mRNA	linear	EST 24-JAN-2002
DEFINITION	BJ171523 full length cDNA library, chloronemata and young			

Accession
R1171523
gametophytes Physcomitrella patens subsp. patens cDNA clone
pph29d18 3', mRNA sequence.

VERSION	BU171523.1	GI:18339496
KEYWORDS	EST.	

SOURCE ORGANISM	Physcomitrella patens subsp. patens	Physcomitrella patens subsp. patens
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

REFERENCE

1 (bases 1 to 692)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrellales

AUTHORS
Fujita, T., Shii-1, T., Seki, M., Kamiya, A., Uchiyama, I., Nishiyama, T.,
, Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe
, M.

TITLE	Comparison of the moss <i>Physcomitrella patens</i> genome with flowering plants genome
JOURNAL	Unpublished (2002)
COMMENT	Contact: Tadasu Shin-i

Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856

Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp
A backbone of the network is

As a substrate, the vector is basically from pRivuescript(Ki), that was in vivo excised from a modified λ phage vector (Mobi T₁, Germany). 3' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA and flanked with conventional T7 and T3 primers. cDNA insert could be amplified with conventional PCR primers. This full-length cDNA library was generated basically according to the method described in the Plant J 15, 707-720 (1998) Seki M. et al. Protoplasts were blended by the POLYTRON, and then cultivated on the BCDAT medium for 13-14 days under the continuous light.

FEATURES
SOURCE

```
/organism="Physcomitrella patens subsp. patens"  
/db_xref="taxon:145481"  
/clone="pph29d18"
```

/clone_11b="full length cDNA library, chloronemata and
young gametophores"
/tissue_type="mixture of chloronemata and young
gametophores with 2 to 5 leaves"
BASE COUNT 213 a 153 c 148 g 178 t
ORIGIN

Query Match 79.1%; Score 17.4; DB 13; Length 692;
Best Local Similarity 94.7%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGACTTCTGAGTACCAAT 19
|||||
Db 337 AAGACTTCTGAGTACCAAT 355

Search completed: March 17, 2003, 13:09:28
JOB time : 782.688 secs